Another aspect of the invention is directed to a computer-implemented method. For instance, a computer and/or an automated system may be provided that is able to automatically and/or repetitively perform any of the methods described herein. As used herein, "automated" devices refer to devices that are able to operate without human direction, i.e., an automated device can perform a function during a period of time after any human has finished taking any action to promote the function, e.g. by entering instructions into a computer to start the process. Typically, automated equipment can perform repetitive functions after this point in time. The processing steps may also be recorded onto a machine-readable medium in some cases.

For example, in some cases, a computer may be used to control imaging of the sample, e.g., using fluorescence microscopy, STORM or other super-resolution techniques such as those described herein. In some cases, the computer may also control operations such as drift correction, physical registration, hybridization and cluster alignment in image analysis, cluster decoding (e.g., fluorescent cluster decoding), error detection or correction (e.g., as discussed herein), noise reduction, identification of foreground features from background features (such as noise or debris in images), or the like. As an example, the computer may be used to control activation and/or excitation of signaling entities within the sample, and/or the acquisition of images of the signaling entities. In one set of embodiments, a sample may be excited using light having various wavelengths and/or intensities, and the sequence of the wavelengths of light used to excite the sample may be correlated, using a computer, to the images acquired of the sample containing the signaling entities. For instance, the computer may apply light having various wavelengths and/or intensities to a sample to yield different average numbers of signaling entities in each region of interest (e.g., one activated entity per location, two activated entities per location, etc.). In some cases, this information may be used to construct an image and/or determine the locations of the signaling entities, in some cases at high resolutions, as noted above.

In some aspects, the sample is positioned on a microscope. In some cases, the microscope may contain one or more channels, such as microfluidic channels, to direct or control fluid to or from the sample. For instance, in one embodiment, nucleic acid probes such as those discussed herein may be introduced and/or removed from the sample by flowing fluid through one or more channels to or from the sample. In some cases, there may also be one or more chambers or reservoirs for holding fluid, e.g., in fluidic communication with the channel, and/or with the sample. Those of ordinary skill in the art will be familiar with channels, including microfluidic channels, for moving fluid to or from a sample.

10x Genomics EX1004 (Part 2)

As used herein, "microfluidic," "microscopic," "microscale," the "micro-" prefix (for example, as in "microchannel"), and the like generally refers to elements or articles having widths or diameters of less than about 1 mm, and less than about 100 microns (micrometers) in some cases. In some embodiments, larger channels may be used instead of, or in conjunction with, microfluidic channels for any of the embodiments discussed herein. For examples, channels having widths or diameters of less than about 10 mm, less than about 9 mm, less than about 8 mm, less than about 7 mm, less than about 6 mm, less than about 5 mm, less than about 4 mm, less than about 3 mm, or less than about 2 mm may be used in certain instances. In some cases, the element or article includes a channel through which a fluid can flow. In all embodiments, specified widths can be a smallest width (i.e. a width as specified where, at that location, the article can have a larger width in a different dimension), or a largest width (i.e. where, at that location, the article has a width that is no wider than as specified, but can have a length that is greater). Thus, for instance, the microfluidic channel may have an average cross-sectional dimension (e.g., perpendicular to the direction of flow of fluid in the microfluidic channel) of less than about 1 mm, less than about 500 microns, less than about 300 microns, or less than about 100 microns. In some cases, the microfluidic channel may have an average diameter of less than about 60 microns, less than about 50 microns, less than about 40 microns, less than about 30 microns, less than about 25 microns, less than about 10 microns, less than about 5 microns, less than about 3 microns, or less than about 1 micron.

A "channel," as used herein, means a feature on or in an article (e.g., a substrate) that at least partially directs the flow of a fluid. In some cases, the channel may be formed, at least in part, by a single component, e.g. an etched substrate or molded unit. The channel can have any cross-sectional shape, for example, circular, oval, triangular, irregular, square or rectangular (having any aspect ratio), or the like, and can be covered or uncovered (i.e., open to the external environment surrounding the channel). In embodiments where the channel is completely covered, at least one portion of the channel can have a cross-section that is completely enclosed, and/or the entire channel may be completely enclosed along its entire length with the exception of its inlet and outlet.

A channel may have any aspect ratio, e.g., an aspect ratio (length to average cross-sectional dimension) of at least about 2:1, more typically at least about 3:1, at least about 5:1, at least about 10:1, etc. As used herein, a "cross-sectional dimension," in reference to a fluidic or microfluidic channel, is measured in a direction generally perpendicular to fluid flow within the channel. A channel generally will include characteristics that facilitate

control over fluid transport, e.g., structural characteristics and/or physical or chemical characteristics (hydrophobicity vs. hydrophilicity) and/or other characteristics that can exert a force (e.g., a containing force) on a fluid. The fluid within the channel may partially or completely fill the channel. In some cases the fluid may be held or confined within the channel or a portion of the channel in some fashion, for example, using surface tension (e.g., such that the fluid is held within the channel within a meniscus, such as a concave or convex meniscus). In an article or substrate, some (or all) of the channels may be of a particular size or less, for example, having a largest dimension perpendicular to fluid flow of less than about 5 mm, less than about 2 mm, less than about 1 mm, less than about 500 microns, less than about 200 microns, less than about 100 microns, less than about 60 microns, less than about 50 microns, less than about 40 microns, less than about 30 microns, less than about 25 microns, less than about 10 microns, less than about 3 microns, less than about 1 micron, less than about 300 nm, less than about 100 nm, less than about 30 nm, or less than about 10 nm or less in some cases. In one embodiment, the channel is a capillary.

A variety of materials and methods, according to certain aspects of the invention, can be used to form devices or components containing microfluidic channels, chambers, etc. For example, various devices or components can be formed from solid materials, in which the channels can be formed via micromachining, film deposition processes such as spin coating and chemical vapor deposition, physical vapor deposition, laser fabrication, photolithographic techniques, etching methods including wet chemical or plasma processes, electrodeposition, and the like. See, for example, *Scientific American*, 248:44-55, 1983 (Angell, *et al*).

In one set of embodiments, various structures or components can be formed of a polymer, for example, an elastomeric polymer such as polydimethylsiloxane ("PDMS"), polytetrafluoroethylene ("PTFE" or Teflon®), or the like. For instance, according to one embodiment, a channel such as a microfluidic channel may be implemented by fabricating the fluidic system separately using PDMS or other soft lithography techniques (details of soft lithography techniques suitable for this embodiment are discussed in the references entitled "Soft Lithography," by Younan Xia and George M. Whitesides, published in the *Annual Review of Material Science*, 1998, Vol. 28, pages 153-184, and "Soft Lithography in Biology and Biochemistry," by George M. Whitesides, Emanuele Ostuni, Shuichi Takayama, Xingyu Jiang and Donald E. Ingber, published in the *Annual Review of Biomedical Engineering*, 2001, Vol. 3, pages 335-373; each of these references is incorporated herein by reference).

Other examples of potentially suitable polymers include, but are not limited to, polyethylene terephthalate (PET), polyacrylate, polymethacrylate, polycarbonate,

polystyrene, polyethylene, polypropylene, polyvinylchloride, cyclic olefin copolymer (COC), polytetrafluoroethylene, a fluorinated polymer, a silicone such as polydimethylsiloxane, polyvinylidene chloride, bis-benzocyclobutene ("BCB"), a polyimide, a fluorinated derivative of a polyimide, or the like. Combinations, copolymers, or blends involving polymers including those described above are also envisioned. The device may also be formed from composite materials, for example, a composite of a polymer and a semiconductor material.

In some embodiments, various microfluidic structures or components of the device are fabricated from polymeric and/or flexible and/or elastomeric materials, and can be conveniently formed of a hardenable fluid, facilitating fabrication via molding (e.g. replica molding, injection molding, cast molding, etc.). The hardenable fluid can be essentially any fluid that can be induced to solidify, or that spontaneously solidifies, into a solid capable of containing and/or transporting fluids contemplated for use in and with the fluidic network. In one embodiment, the hardenable fluid comprises a polymeric liquid or a liquid polymeric precursor (i.e. a "prepolymer"). Suitable polymeric liquids can include, for example, thermoplastic polymers, thermoset polymers, waxes, metals, or mixtures or composites thereof heated above their melting point. As another example, a suitable polymeric liquid may include a solution of one or more polymers in a suitable solvent, which solution forms a solid polymeric material upon removal of the solvent, for example, by evaporation. Such polymeric materials, which can be solidified from, for example, a melt state or by solvent evaporation, are well known to those of ordinary skill in the art. A variety of polymeric materials, many of which are elastomeric, are suitable, and are also suitable for forming molds or mold masters, for embodiments where one or both of the mold masters is composed of an elastomeric material. A non-limiting list of examples of such polymers includes polymers of the general classes of silicone polymers, epoxy polymers, and acrylate polymers. Epoxy polymers are characterized by the presence of a three-membered cyclic ether group commonly referred to as an epoxy group, 1,2-epoxide, or oxirane. For example, diglycidyl ethers of bisphenol A can be used, in addition to compounds based on aromatic amine, triazine, and cycloaliphatic backbones. Another example includes the well-known Novolac polymers. Non-limiting examples of silicone elastomers suitable for use according to the invention include those formed from precursors including the chlorosilanes such as methylchlorosilanes, ethylchlorosilanes, phenylchlorosilanes, etc.

Silicone polymers are used in certain embodiments, for example, the silicone elastomer polydimethylsiloxane. Non-limiting examples of PDMS polymers include those

sold under the trademark Sylgard by Dow Chemical Co., Midland, MI, and particularly Sylgard 182, Sylgard 184, and Sylgard 186. Silicone polymers including PDMS have several beneficial properties simplifying fabrication of various structures of the invention. For instance, such materials are inexpensive, readily available, and can be solidified from a prepolymeric liquid via curing with heat. For example, PDMSs are typically curable by exposure of the prepolymeric liquid to temperatures of about, for example, about 65 °C to about 75 °C for exposure times of, for example, at least about an hour. Also, silicone polymers, such as PDMS, can be elastomeric and thus may be useful for forming very small features with relatively high aspect ratios, necessary in certain embodiments of the invention. Flexible (e.g., elastomeric) molds or masters can be advantageous in this regard.

One advantage of forming structures such as microfluidic structures or channels from silicone polymers, such as PDMS, is the ability of such polymers to be oxidized, for example by exposure to an oxygen-containing plasma such as an air plasma, so that the oxidized structures contain, at their surface, chemical groups capable of cross-linking to other oxidized silicone polymer surfaces or to the oxidized surfaces of a variety of other polymeric and non-polymeric materials. Thus, structures can be fabricated and then oxidized and essentially irreversibly sealed to other silicone polymer surfaces, or to the surfaces of other substrates reactive with the oxidized silicone polymer surfaces, without the need for separate adhesives or other sealing means. In most cases, sealing can be completed simply by contacting an oxidized silicone surface to another surface without the need to apply auxiliary pressure to form the seal. That is, the pre-oxidized silicone surface acts as a contact adhesive against suitable mating surfaces. Specifically, in addition to being irreversibly sealable to itself, oxidized silicone such as oxidized PDMS can also be sealed irreversibly to a range of oxidized materials other than itself including, for example, glass, silicon, silicon oxide, quartz, silicon nitride, polyethylene, polystyrene, glassy carbon, and epoxy polymers, which have been oxidized in a similar fashion to the PDMS surface (for example, via exposure to an oxygen-containing plasma). Oxidation and sealing methods useful in the context of the present invention, as well as overall molding techniques, are described in the art, for example, in an article entitled "Rapid Prototyping of Microfluidic Systems and Polydimethylsiloxane," *Anal. Chem.*, 70:474-480, 1998 (Duffy *et al.*), incorporated herein by reference.

The following documents are each incorporated herein by reference in their entireties: U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three

Dimensions," by Zhuang, *et al.*; and Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*

In addition, incorporated herein by reference in their entireties are U.S. Provisional Patent Application Serial No. 62/031,062, filed July 30, 2014, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/050,636, filed September 15, 2014, entitled "Probe Library Construction," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/142,653, filed April 3, 2015, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; and a PCT application filed on even date herewith, entitled "Probe Library Construction," by Zhuang, *et al.*

The following examples are intended to illustrate certain embodiments of the present invention, but do not exemplify the full scope of the invention.

EXAMPLE 1

The example presents a platform to enable the simultaneous detection of the number and spatial organization of thousands of distinct mRNAs within single cells with high efficiency and low error-rate using a novel form of highly multiplexed fluorescence *in situ* hybridization (FISH). This example accomplishes these measurements by integrating and innovating methods for massively parallel probe synthesis, super-resolution imaging, and self-correcting error-checking codes.

Here, these examples present methods for the simultaneous detection of some or all of the thousands of unique RNAs expressed in a cell. This approach not only promises to revolutionize the throughput of the already effective single-molecule FISH (smFISH) approach, but also allows researchers to benefit from the hypothesis free discovery approach which has made other whole-genome systems approaches to biology so effective. For example, this whole genome approach may allow researchers to discover RNAs whose expression levels and/or subcellular localization patterns change under certain conditions of interest, such as disease states, without knowing, *a priori*, which mRNA will change in abundance or localization. Simultaneous measurements of hundreds of genes within a single cell also allow for the identification of correlations between genes in expression and localization patterns in some cases.

This can be achieved using methods for highly multiplexed smFISH via the sequential hybridization of orthogonal detection probes and super-resolution imaging, reducing the cost of probe synthesis, and the development of a highly automated system to minimize demands on the user, as discussed herein. This provides an integrated platform to handle the

bioinformatics of probe design, the mathematics of error-correcting codes, the complexity of image registration and analysis, and the cumbersome fluid handling through a simple suite of user friendly interfaces. This integration allows easy operation with limited user training and facilitates the rapid collection of data.

This example illustrates: (1) computational design of "codewords" to attach to all RNA targets in the cell that will allow unique identification of each RNA with some degree of experimental error tolerance, (2) translation of these codewords into nucleotide sequences and synthesis of the required single-stranded (ss) oligonucleotide (e.g. ssDNA) probes, (3) sample fixation and hybridization of these probes to the RNA targets *in situ*, (4) read-out of these codewords via successive rounds of hybridization of distinct fluorescent probes imaged with conventional fluorescence microscopy or super-resolution fluorescence microscopy, and (5) automated decoding of measured codewords combined with computational error correction to uniquely and robustly identify individual mRNAs.

In the first step, a "codeword" is assigned to every RNA that is to be labeled. In a typical design these may be strings of N binary letters or positions. Codewords may be chosen from the same wide range of existing error tolerant or error-correcting encoding schemes developed for digital storage and communication, e.g., using Hamming codes or the like. For example, actin-RNA may be assigned the binary codeword 11001010. Each codeword may be unique and separated from the other codewords by a Hamming distance, h, which measures the number of letters or positions that must be incorrectly read for one codeword to be misinterpreted as a different one. A Hamming distance greater than 1 between all codewords allows for some measurement errors to be detected—since simple errors would produce codewords that are not used to encode RNAs. For a Hamming distance larger than 2, it is also possible to correct some errors, as codewords with one error will be closest in Hamming distance to a single, unique codeword. The total number of different RNAs to be detected from the transcriptome and the amount of error correction desired determines the length of the codewords. Information theory provides several efficient algorithms for assembling error-correcting binary codebooks.

In the second step, this encoding scheme is translated into a set of oligonucleotide (e.g. DNA) probe sequences, which can be called primary probes or encoding probes, each of which not only targets a probe to the RNA of interest but also encodes the unique binary codeword within a set of secondary binding sites (Fig. 1). For example, first designed may be primary binding sequences for each targeted mRNA. These sequences are "target sequences" that are comprised of complementary nucleotide sequences to their target RNAs

computationally selected to satisfy a stringent set of hybridization conditions, including uniqueness in the target genome. To improve the efficiency of hybridization to individual mRNAs, multiple primary target sequences are designed for each individual RNA. Then each position within the set of codewords is assigned a unique oligonucleotide (e.g. DNA) sequence, which is called a read sequence. These tags are designed as to have no interaction with endogenous mRNA sequences or each other. For instance, for all the value "1"s in a codeword of an individual mRNA, the corresponding read sequence is attached to the primary targeting sequences against that mRNA. In general, each probe will contain a target sequence and one or more read sequences. If the total length of the necessary read sequences and the primary target sequence exceeds synthesis capabilities, then subsets of the read sequences can be appended to distinct target sequences. For example, consider the potential codeword 11001010 for actin. Probe sequences for this RNA could contain the read sequences corresponding to positions 1, 2, 5, and 7 in the codeword attached to a variety of target sequences specific to actin. After all the sequences have been designed, the resulting complex set of unique custom oligonucleotide (e.g. DNA) sequences is manufactured and amplified using methods as described below.

In the third step, the resulting pool of DNA is hybridized, e.g., to fixed, permeabilized cells. In this process, individual probes may be attached to every RNA in the cell by hybridization of their corresponding target sequences with the RNA while the read sequences remain free to bind the appropriate secondary probes as discussed below.

In the fourth step—the read-out step—fluorescently labeled secondary nucleic acid probes (also called readout probes) are successively hybridized to the read sequences attached to the target sequences that binds to the mRNA targets in the above step. When simultaneously imaging a large number of different RNA species in cells, the density of labeled RNAs may exceed that at which each RNA can be resolved via conventional imaging methods. Thus, this may be performed using a super-resolution imaging method, for example STORM (stochastic optical reconstruction microscopy), to resolve the labeled molecules. After each round of hybridization and imaging with the secondary probes, the fluorophores are quenched or otherwise inactivated either via chemical or optical techniques such as oxidation, chemically bleaching, photobleaching, stringent washing or enzymatic digestion, etc. The sample is then stained with the next secondary probe, and the cycle continues until all positions of the codewords have been read out. In the simplest incarnation, there will be one hybridization step for each position within the codeword, e.g. 8 hybridization steps for an 8-letter codeword (Fig. 1).

Fig. 1 shows schematic diagrams of this example. Fig. 1A shows that every position of the codewords is assigned a unique oligonucleotide sequence when this position has a value "1." All mRNA codewords are then translated into combinations of read sequences, which are attached to the targeting sequence. Fig. 1B shows various steps of the labeling scheme of this example. In the first step, all mRNAs (I - III) are tagged with multiple oligonucleotide (e.g. ssDNA) probes comprising a primary targeting sequence, which hybridizes to the RNA of interest, and a "tail" (i.e. containing read sequences) carrying the translated codeword, which does not interact with endogenous nucleotide sequences. In the next step, the first secondary probe is added, which can bind all probes whose tails have a read sequence corresponding to the value of "1" in the first position. The dyes on these secondary probes are imaged and bleached, then the next secondary probe is added to bind probes attached to mRNA which have a value of "1" in the second position of their assigned codeword, and so on.

In the final step, the microscopy images from each staining and imaging round are aligned, for example, computationally (e.g. using fiducial beads or other markers tracked during image acquisition), and the clusters of localizations resolved by conventional fluorescence microscopy or super-resolution imaging (e.g. STORM) from the different rounds are identified. These clusters of localizations arise from individual target mRNA molecules, and the hybridization rounds in which a spot was detected in a given cluster correspond to the "1" in the codeword for that mRNA. If there are no missed-detection events or false positive signal in the images, this codeword will perfectly match one of the expected codewords. Fig. 1 describes an example in which the codeword has three letters, i.e. three positions, and the three target mRNAs have codewords 110, 101, and 011 assigned to them. In real experimental examples, the codeword could contain more digits. For example, the mRNA for actin can be assigned the codeword, 11001010. In that case, detected clusters containing overlapping localization signals in the 1st, 2nd, 5th and 7th hybridization steps (meaning the 1st, 2nd, 5th and 7th secondary probes bound to this site) can be identified as individual actin mRNA molecules, since the pattern of positive bindings matches the codewode of actin (11001010). In addition, if there are missed-detection events or false positive signals in the image data, these aberrations can be corrected by the implemented error-correction scheme. For example, clusters of localizations with a detected codeword that has only one digit discrepancy from 11001010 (such as 11000010 or 11101010) can also be identified as actin mRNA since all other valid codeword in this example differ from the detected pattern in two more positions.

EXAMPLE 2

This example describes another alternative approach that differs in several of the steps described above. This approach begins with the first step, construction of the codewords to the desired mRNA targets, as described above.

In the second step of this approach, nucleic acid probes are designed that bind uniquely to the mRNA targets of interest, as described above. However, instead of appending unique read sequences to these targeting sequences, unique pools or groups of probes are constructed from these target sequences. Each pool comprises all or a subset of the sequences that target all mRNAs which contain the same value at a given position in their codeword. For example, the first pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the first position of their codewords, e.g. 110 and 101 but not 011; the second pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the second position of their codewords, e.g. 110 and 011 but not 101; the third pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the third position of their codewords, e.g. 011 and 101 but not 110 (Fig. 1C). As another example, consider the potential codeword 11001010 for actin. Probes that target this mRNA would be included in pools 1, 2, 5, and 7 but not in pools 3, 4, 6, and 8. The same target to a given mRNA may or may not be included in pools. For example, a probe that targets the same region of actin may be included in pools 1, 2, 5, and 7 or any subset of these pools. After all of the pools have been designed, each complex set of unique, custom oligonucleotide sequences is manufactured and amplified using methods as described below.

In the third step of this approach, the first pool of probes is hybridized, e.g., to fixed, permeabilized cells. In this process, the fluorophores attached to each of the probes in this pool are bound to each of the targets of that pool. The binding of these probes is then determined by fluorescence microscopy. As described above, these images can be collected either via a range of methods including both conventional fluorescence imaging or super-resolution imaging methods such as STORM. After a round of imaging, the probes from the first pool are inactivated or removed from the sample via the methods described above. This process is then repeated for each successive pool of probes until some or all of the pools have been applied to the sample and imaged such that all positions in the codewords have been read out. In the simplest incarnation, there will be one hybridization and imaging step for each position in the codeword, e.g. 3 rounds of hybridization and imaging for a codeword with 3 positions (Fig 1C) or 8 rounds of hybridization and imaging for a codeword with 8 positions.

The final step of this approach is identical to that described above.

EXAMPLE 3

In this example, 14 genes (PGK1, H3F3B, PKM, ENO1, GPI, EEF2, GNAS, HSPA8, GAPDH, CALM1, RHOA, PPIA, UBA52, and VCP) were encoded using a subset of the (8,4) SECDED code (Figs. 2A-2E). To determine the accuracy of these measurements, the measured abundances of these 14 mRNAs were compared to the abundances measured from bulk RNA-seq of A549 cells (published ENCODE data). Remarkably, it was found that there was excellent agreement between these two measurements, as the transcripts count measured using the sequential hybridization approach correlated with gene expression measured using RNA-seq with a Pearson correlation coefficient r of 0.75 (Fig. 2F). Gene expression from 3 other cells were also measured, and it was found that the gene expression of these 14 genes was highly correlated among the cells with a r of 0.96 (Fig. 2G).

Codebook Design. Each mRNA in the target set was assigned a binary codeword using a Single Error Correction Double Error Detection (SECDED) code. SECDED is an extended Hamming codebook with an additional parity bit. Briefly, Matlab's Communications System toolbox was used to generate SECDED codes of either 8 or 16 letters or positions. In both cases, only those codewords containing four 1s were used. These words were assigned at random to mRNAs in the target set. [0 1 0 1 1 1 0 0] is an example of the 8-letter codewords used (i.e., these codewords each contained four 1s and four 0s.) [0 1 0 1 1 1 0 0 0 0 0 0 0 0 0 0] is an example of the 16-letter codewords used (i.e., each codeword contained four 1s and twelve 0s). Not every codeword was necessarily assigned to an mRNA.

Computational Assembly of ssDNA Primary Probe Sequence. The number of primary nucleic acid probes used for hybridization with mRNA targets ranged from 200 to 2000 unique oligonucelotides, depending on the experiment. For example, to label 14 mRNAs with 28 oligos targeting each gene, 392 unique sequences were used. Large number of oligos with unique sequences were purchased in a pool from LC Sciences or CustomArray. However, array synthesized oligos were in minute quantity that was insufficient for *in situ* hybridization. The protocol for their amplification are described below.

Each primary probe contained three components: flanking primer sequences to allow enzymatic amplification of probes, targeting sequence for *in situ* hybridization to mRNAs, and secondary tag sequence containing one or more read sequences for sequential readout of codewords.

The following is an example of a primary probe:

GTTGGCGACGAAAGCACTGCGATTGGAACCGTCCCAAGCGTTGCG
CTTAATGGATCATCAATTTTGTCTCACTACGACGGTCAATCGCGCTGCATACTTG
CGTCGGTCGGACAAACGAGG

(SEQ ID NO: 1)

The components are arranged in the following order: forward primer (not underlined), secondary read sequence 1 (underlined), mRNA targeting sequence (not underlined), secondary read sequence 2 (underlined), and reverse primer (not underlined). The secondary read sequences are the reverse complement of the corresponding secondary probes. Since only codewords that contained four '1's were used, the primary probes for each mRNA needed to contain 4 different secondary read sequences in this example. However, in order to reduce the overall length of the primary probes, the pool of targeting sequences for each mRNA target was split at random into two pools. Two secondary read sequences are attached each probe in one of the two pools and the other two secondary read sequences are attached the probes in the other pool. The design criteria for each component are described below.

Primer Design. Specific index primers were generated by a collection of 240,000 published sequences of orthogonal 25-bp long sequences. These sequences were trimmed to 20 bp, selected for a narrow 70 to 80 ^{o}C melting temperature, the absence of consecutive repeats of 3 or more base, and the presence of a GC clamp, i.e. one of the two 3' terminal bases must be G or C. To further improve specificity, these sequences were then screened against the human genome using BLAST+ (Camacho et al 2009), and primers with 14 or more contiguous bases of homology were eliminated. In a subsequent screening via BLAST+, primers that shared 11 or more contiguous bases or more than 5 bases at the 3' end of any other primer or the T7 promoter were also removed.

Secondary Probes Design. 30-bp long secondary probe sequences were created by concatenating fragments of the orthogonal primer set described above. These secondaries were then screened for orthogonality with other secondaries (no more than 11 basepairs of homology) and for potential off-target binding sites in the human genome (no more than 14 basepairs of homology). Secondary sequences used in this example are provided in Table 1.

Table 1

Bit	Secondary sequences	Sequence number
B1	CGCAACGCTTGGGACGGTTCCAATCGGATC	SEQ ID NO: 2
B2	CGAATGCTCTGGCCTCGAACGAACGATAGC	SEQ ID NO: 3
B3	ACAAATCCGACCAGATCGGACGATCATGGG	SEQ ID NO: 4
B4	CAAGTATGCAGCGCGATTGACCGTCTCGTT	SEQ ID NO: 5
B5	TGCGTCGTCTGGCTAGCACGGCACGCAAAT	SEQ ID NO: 6
B6	AAGTCGTACGCCGATGCGCAGCAATTCACT	SEQ ID NO: 7
B7	CGAAACATCGGCCACGGTCCCGTTGAACTT	SEQ ID NO: 8
B8	ACGAATCCACCGTCCAGCGCGTCAAACAGA	SEQ ID NO: 9
B9	CGCGAAATCCCCGTAACGAGCGTCCCTTGC	SEQ ID NO: 10
B10	GCATGAGTTGCCTGGCGTTGCGACGACTAA	SEQ ID NO: 11
B11	CCGTCGTCTCCGGTCCACCGTTGCGCTTAC	SEQ ID NO: 12
B12	GGCCAATGGCCCAGGTCCGTCACGCAATTT	SEQ ID NO: 13
B13	TTGATCGAATCGGAGCGTAGCGGAATCTGC	SEQ ID NO: 14
B14	CGCGCGGATCCGCTTGTCGGGAACGGATAC	SEQ ID NO: 15
B15	GCCTCGATTACGACGGATGTAATTCGGCCG	SEQ ID NO: 16
B16	GCCCGTATTCCCGCTTGCGAGTAGGGCAAT	SEQ ID NO: 17

mRNA Targeting Sequence Design. To determine the relative abundance of all the isoforms of all genes expressed in these cell lines, transcriptome profiling data from the ENCODE project for total RNA from A549 and IMR90 cells was processed using the publicly available software *cufflinks*, along with human genome annotations from gencode v18. Gene models corresponding to the most highly expressed isoform were used to build a sequence library in FASTA format recording the dominant isoform of every gene. Genes of interest were selected from this library. These genes were partitioned into 1kb segments, then the software *OligoArray2.1* was used to generate primary probe sequences for the human transcriptome with the following constraints: 30-bp or 40-bp length, depending on the experiment; probe-target melting temperatures greater than 70 °C (variable parameter); no cross hybridization targets with melting temperatures greater than 72 °C (variable parameter); no predicted internal secondary structures with melting temperatures greater than 76 °C (variable parameter); and no single-nucleotide contiguous repeats of 6 or more bases. After OligoArray probe selection, all potential probes that mapped to a different gene were rejected

while all potential probes with multiple alignments to the same gene were retained. A BLAST database was assembled from the FASTA library of all expressed genes to screen for probes' uniqueness. For each gene, 14 to 28 targeting sequences produced during the OligoArray processing were selected.

Probe Synthesis – index PCR. The template for specific probe sets were selected from the complex oligopool via limited-cycle PCR. Briefly, 0.5 to1 ng of the complex oligopool was combined with 0.5 micromolar of each primer. The forward primer matched the priming sequence for the desired subset while the reverse primer was a 5' concatenation of this sequence with a T7 promoter. To avoid the generation of G-quadruplets, which can be difficult to synthesize, the terminal Gs required in the T7 promoter were generated from Gs located at the 5' of the priming region where appropriate. All primers were synthesized by IDT. A 50 microliter reaction volume was amplified either using the KAPA real-time library amplification kit (KAPA Biosystems; KK2701) or via a homemade qPCR mix which included 0.8X EvaGreen (Biotum; 31000-T) and the hot-start Phusion polymerase (New England Biolabs; M0535S). Amplification was followed in real time using Agilent's MX300P or Biorad's CFX Connect. Individual samples were removed immediately before the plateau in amplification to minimize distortion of template abundance due to over-amplification. Individual templates were purified with columns according to the manufacturer's instructions (Zymo DNA Clean and Concentrator; D4003) and eluted in RNase-free deionized water.

Amplification via *in-vitro* transcription. The template was then amplified via *in vitro* transcription. Briefly, 0.5 to 1 microgram of template DNA was amplified into 100-200 microgram of RNA in a single 20-30 microliter reaction with a high yield RNA polymerase (New England Biolabs; E2040S). Reactions were supplemented with 1X RNase inhibitor (Promega RNasin; N2611). Amplification was typically run for 4 to 16 hours at 37 ^{o}C to maximize the yield. The RNA was not purified after the reaction and was either stored at -80 ^{o}C or immediately converted into DNA as described below.

Reverse Transcription. 1 to 2 nmol of fluorescently-labeled ssDNA probe was created from the above *in vitro* transcription reactions using the reverse transcriptase Maxima H- (Thermo Scientific; EP0751). This enzyme was used because of its higher processivity and temperature resistance, which allowed the conversion of large quantities of RNA into DNA within small volumes at temperatures that disfavor secondary structure formation. The unpurified RNA created above was supplemented with 1.6 mM of each dNTP, 1-2 nmol of fluorescently labeled forward primer, 300 units of Maxima H-, 60 units of RNasin, and a final

1X concentration of the Maxima RT buffer. The final 75 microliter volume was incubated at 50 °C for 60 minutes.

Strand Selection and Purification. The template RNA in the reaction above was then removed from the DNA via alkaline hydrolysis. 75 microliters of 0.25 M EDTA and 0.5 N NaOH were added to each reverse transcription reaction, and the sample was incubated at 95 °C for 10 minutes. The reaction was immediately neutralized by purifying the ssDNA probe with a modified version of the Zymo Oligo Clean and Concentrator protocol. Specifically, the 5-microgram capacity column was replaced with a 25-microgram or 100-microgram capacity DNA column as appropriate. The remainder of the protocol was run according to the manufacturer's instructions. Probe was eluted in 100 microliter RNase-free deionized water and evaporated in a vacuum concentrator. The final pellet was resuspended in 10 microliters of RNase-free water and stored at -20 °C. Denaturing poly-acrylimide gel electrophoresis and absorption spectroscopy revealed that this protocol typically produced 90-100% incorporation of the fluorescent primer into full length probe and 75-90% recovery of the total fluorescent probe. Thus, without exceeding a 150-microliter reaction volume, this protocol can be used to create ~2 nmol of fluorescent probe.

Cell culture and fixation. A549 and IMR90 cells (American Type Culture Collection) were cultured with Dulbecco's Modified Eagle Medium and Eagle's Minimum Essential Medium respectively. Cells were incubated at 37 °C with 5% CO_2 for 36-48 hours. Cells were fixed in 3% paraformaldehyde (Electron Microscopy Sciences) in PBS for 15 minutes, washed with PBS, and permeabilized in 70% ethanol overnight at 4 °C.

Fluorescence In Situ Hybridization (FISH) – primary (encoding) probes. Cells were hydrated in wash buffers (2xSSC, 50% formamide) for 10 minutes, labelled with primary oligos (0.5 nM per sequence) in hybridization buffers (2xSSC, 50% formamide, 1 mg/mL yeast tRNA, and 10% dextran sulfate) overnight at 37 °C, washed with wash buffers at 47 °C for 10 minutes twice, and washed with 2xSSC twice. Fluorescent fiducial beads (Molecular Probes, F-8809) were added at a 1:10,000 dilution in 2xSSC before imaging.

Secondary probes. Secondary (readout) probes (10 nM) were hybridized in secondary hybridization buffers (2xSSC, 20% formamide, and 10% dextran sulfate) to their primary targets for 30 minutes at 37 °C. Cells remained on the microscope stage during the hybridization. An objective heater was used to maintain the temperature at 37 °C. Cells were washed with secondary wash buffers (2xSSC, 20% formamide).

Fluidics and STORM Imaging. Multiple rounds of sequential labeling, washing, imaging, and bleaching were performed on an automated platform consisted of a fluidics

setup and a STORM (stochastic optical reconstruction microscopy) microscope. The fluidics setup included a flow chamber (Bioptech FCS2), a peristaltic pump (Rainin Dynamax RP-1), and three computer-controlled 8-way valves (Hamilton MVP and Hamilton HVXM 8-5). This system allowed the automated integration of STORM movie collection and secondary hybridization.

The imaging buffer included, 50 mM Tris (pH 8) 10% (w/v) glucose, 1% βME (2-mercaptoethanol) or 25 mM MEA, with or without 2 mM 1,5-cyclooctadiene, and an oxygen scavenging system (0.5 mg/ml glucose oxidase (Sigma-Aldrich) and 40 microgram/ml catalase (Sigma-Aldrich)). A layer of mineral oil was used to seal the imaging buffer, preventing its acidification over the course of multiple hybridizations.

The STORM setup included an Olympus IX-71 inverted microscope configured for oblique incidence excitation. The samples were continuously illuminated with a 642-nm diode-pumped solid-state laser (VFL-P500-642; MPB communications). A 405-nm solid-state laser (Cube 405-100C; Coherent) was used for activation of dyes. Fluorescence was collected using an Olympus (UPlanSApo 100x, 1.4 NA) objective lens and passed through a custom dichroic, as well as a quad-view beam splitter. All movies were recorded using an EMCCD camera (Andor iAxon 897), imaging at 60Hz. The 512x256 field of view of the camera was split into separate 256x256 pixel movies prior to saving. The left half of this field of view contained the STORM data and the right half contained images of the fluorescent feducial beads. These latter movies were downsampled to 1 Hz prior to saving. During data acquisition, a home-built focus lock was used to maintain a constant focal plane. STORM movies included 20,000 to 30,000 frames in STORM buffer while the bleach movies included 10,000 frames in wash buffer.

Image Analysis – analysis of single-molecule localizations. Movies of single molecule localizations and fluorescent feducial beads were processed separately using a previously published single-emitter localization software.

Image Registration. The starting position of the beads from each round of hybridization were used to align movies from each round. The 2D autocorrelation between bead images of consecutive hybridizations followed by nearest-neighbor matching was used to match beads between images. The pair of beads with the most similar displacement vector were used to compute a rigid translation-rotation warp to align the beads. This alignment method is robust to samples in which multiple feducials are displaced or come detached and reattach during imaging.

Drift Correction. Drift during image acquisition was corrected using the trajectory of the feducial beads (recorded at 1 Hz). Bead positions were linked in each frame. The trajectory of the two beads that moved in the most correlated fashion was taken as the drift trajectory.

mRNA Cluster Calling. Localizations were first screened to be above a threshold number of photons (generally 2000) and required to be within 32 nm of 5 other localizations (parameters may be tuned). The remaining molecule localizations were binned in a 2D histogram of 10x10 nm bins (bin size is a variable parameter). All connected bins were taken to be part of a cluster (diagonal contacts are classified as connected). Clusters were required to have more than 80 total localizations across all hybridizations (variable parameter) to be called an mRNA cluster. The weighted centroids of these clusters from the 2D histogram were recorded as the mRNA positions.

A given cluster is recorded to be represented in an individual hybridization round if more than 9 localizations (variable parameter) are found within a 48 nm radius (variable parameter) of the centroid for that mRNA in each hybridization round.

Cluster Decoding. For each mRNA cluster, a codeword is readout, including "0"s for all the hybridization rounds in which less than the threshold number of localizations are found near the centroid and "1" s for the rounds where above threshold number of localizations are counted. The SECDED codebook decoded these as either perfect matches to target mRNA codewords, correctable errors which can be unambiguously mapped back to target mRNA, or uncorrectable errors, which differed by two or more letters from the words in the codebook.

Fig. 2A shows a STORM image of a cell. Fig. 2B shows a zoom in of the boxed region in Fig. 2A. Each dot indicates a localization. Localizations from different rounds of imaging are shown differently. Fig. 2C shows a representative cluster of localizations from the boxed region in Fig. 2B. The cluster shows localization signals from 4 different hybridizations. This cluster is a putative mRNA encoded with codeword [0 1 0 1 1 1 0 0]. Fig. 2D shows a reconstructed cell image of 14 genes after decoding and error correction. Each gene is shown differently. Fig. 2E shows measured gene expression for the 14 genes from the cell. Fig. 2F shows a comparison of transcript count with ensemble RNA-sequencing data. Fig. 2G shows correlation of transcript expression level between two cells detected using the described approach.

EXAMPLE 4

The following examples are generally directed to multiplexed single-molecule imaging with error-robust encoding allowing for simultaneous measurements of thousands of RNA species in single cells. In general, knowledge of the expression profile and spatial landscape of RNAs in individual cells is essential for understanding the rich repertoire of cellular behaviors. The following examples reports various techniques directed to single-molecule imaging approaches that allow the copy numbers and spatial localizations of thousands of RNA species to be determined in single cells. Some of these techniques are called Multiplexed Error-Robust Fluorescence in Situ Hybridization or "MERFISH."

Using error-robust encoding schemes to combat single-molecule labeling and detection errors, these examples demonstrated the imaging of hundreds to thousands of unique RNA species in hundreds of individual cells. Correlation analysis of the ~10^4 to ~10^6 pairs of genes allowed constraints on gene regulatory networks, prediction of novel functions for many unannotated genes, and identification of distinct spatial distribution patterns of RNAs that correlate with properties of the encoded proteins.

System-wide analyses of the abundance and spatial organization of RNAs in single cells promise to transform understanding in many areas of cell and developmental biology, such as the mechanisms of gene regulation, the heterogeneous behavior of cells, and the development and maintenance of cell fate. Single-molecule fluorescence *in situ* hybridization (smFISH) has emerged as a powerful tool for studying the copy number and spatial organization of RNAs in single cells either in isolation or in their native tissue context. Taking advantage of its ability to map the spatial distributions of specific RNAs with high resolution, smFISH has revealed the importance of subcellular RNA localization in diverse processes such as cell migration, development, and polarization. In parallel, the ability of smFISH to precisely measure the copy numbers of specific RNAs without amplification bias has allowed quantitative measurement of the natural fluctuations in gene expression, which has in turn elucidated the regulatory mechanisms that shape such fluctuations and their role in a variety of biological processes.

However, application of the smFISH approach to many systems-level questions remains limited by the number of RNA species that can be simultaneously measured in single cells. State-of-the-art efforts using combinatorial labeling by either color-based barcodes or sequential hybridization have enabled simultaneous measurements of 10-30 different RNA species in individual cells, yet many interesting biological questions would benefit from the measurement of hundreds to thousands of RNAs within a single cell, which are not

achievable using such techniques. For example, analysis of how the expression profile of such a large number of RNAs vary from cell to cell and how these variations correlate among different genes could be used to systematically identify co-regulated genes and map regulatory networks; knowledge of the subcellular organizations of numerous RNAs and their correlations could help elucidate molecular mechanisms underlying the establishment and maintenance of many local cellular structures; and RNA profiling of individual cells in native tissues could allow *in situ* identification of cell type.

The following examples generally discuss certain techniques called MERFISH, which are highly multiplexed smFISH imaging methods that substantially increase the number of RNA species that can be simultaneously imaged in single cells by using combinatorial labeling and sequential imaging with error-robust encoding schemes. These examples demonstrate this multiplexed imaging approach by simultaneously measuring 140 RNA species using an encoding scheme that can both detect and correct errors and 1001 RNA species using an encoding scheme that can detect but not correct errors. It should be understood that these numbers are by way of exemplification only, not limitation. Correlation analyses of the copy number variations and spatial distributions of these genes allowed us to identify groups of genes that are co-regulated and groups of genes that share similar spatial distribution patterns inside the cell.

Combinatorial labeling with error-robust encoding schemes. Combinatorial labeling that identifies each RNA species by multiple (N) distinct signals offers a route to rapidly increase the number of RNA species that can be probed simultaneously in individual cells (Fig. 5A). However, this approach to scaling up the throughput of smFISH to the systems scale faces a significant challenge because not only does the number of addressable RNA species increases exponentially with N, but the detection error rates also grow exponentially with N (Figs. 5B-5D). Imagine a conceptually simple scheme to implement combinatorial labeling, where each RNA species is encoded with a N-bit binary word and the sample is probed with N corresponding rounds of hybridization, each round targeting only the subset of RNAs that should read '1' in the corresponding bit (Fig. 11). N rounds of hybridization would allow 2^N-1 RNA species to be probed. With just 16 hybridizations, over 64,000 RNA species, which should cover the entire human transcriptome including both messenger RNAs (mRNAs) and non-coding RNAs, could be identified (Fig. 5B; upper symbols). However, as N increases, the fraction of RNAs properly detected (the calling rate) would rapidly decrease and, more troublingly, the fraction of RNAs that are identified as incorrect species (the misidentification rate) would rapidly increase (Fig. 5C, lower symbols; Fig. 5D, upper

symbols). With realistic error rates per hybridization (measured below), the majority of RNA molecules would be misidentified after 16 rounds of hybridizations!

To address this challenge, error-robust encoding schemes were designed, in which only a subset of the 2^N-1 words separated by a certain Hamming distance were used to encode RNAs. In a codebook where the minimum Hamming distance is 4 (HD4 code), at least four bits must be read incorrectly to change one code word into another (Fig. 12A). As a result, every single-bit error produces a word that is uniquely close to a single code word, allowing such errors to be detected and corrected (Fig. 12B). Double-bit errors produce words with an equal Hamming distance of 2 from multiple code words and, thus, can be detected but not corrected (Fig. 12C). Such a code should substantially increase the calling rate and reduce the misidentification rate (Fig. 5C and 5D, middle symbols). To further account for the fact that it is more likely to miss a hybridization event (an 1-->0 error) than to misidentify a background spot as an RNA (an 0-->1 error) in smFISH measurements, a modified HD4 (MHD4) code was designed, in which the number of '1' bits were kept both constant and relatively low, only four per word, to reduce error and avoid biased detection. This MHD4 code should further increase the calling rate and reduce the misidentification rate (Fig. 5C, upper symbols; Fig. 5D, lower symbols).

In addition to the error considerations, several practical challenges have also made it difficult to probe a large number of RNA species, such as the high cost of the massive number of fluorescently labeled FISH probes needed and the long time required to complete many rounds of hybridization. To overcome these challenges, in this example, a two-step labeling scheme was designed to encode and readout cellular RNAs (Fig. 5E). First, cellular RNAs were labeled with a set of encoding probes (also called primary probes), each probe comprising a RNA targeting sequence and two flanking readout sequences. Four of the *N* unique readout sequences were assigned to each RNA species based on the MHD4 code word of the RNA. Second, these *N* readout sequences were identified with complementary FISH probes, the readout probes (also called secondary probes) via *N* rounds of hybridization and imaging, each round using a unique readout probe. To increase the signal to background ratio, every cellular RNA was labeled with ~192 encoding probes. Because each encoding probe contained two of the four readout sequences associated with that RNA (Fig. 5E), a maximum of ~96 readout probes can bind to each cellular RNA per hybridization round. To generate the massive number of encoding probes required, they were amplified from array-derived oligonucleotide pools containing tens of thousands of custom sequences using an enzymatic amplification process comprising *in vitro* transcription followed by reverse

transcription (Fig. 13, see below regarding probe synthesis). This two-step labeling approach significantly diminished the total hybridization time for an experiment: it was found that efficient hybridization to the readout sequences took only 15 minutes whereas efficient direct hybridization to cellular RNA required more than 10 hours.

Fig. 5 describes MERFISH, a highly multiplexed smFISH approach using combinatorial labeling and error-robust encoding. Fig. 5A shows a schematic depiction of the identification of multiple RNA species in *N* rounds of imaging. Each RNA species is encoded with a *N*-bit binary word and during each round of imaging, only the subset of RNAs that should read '1' in the corresponding bit emit signal. Figs. 5B-5D show the number of addressable RNA species (Fig. 5B), the rate at which these RNAs are properly identified (calling rate) (Fig. 5C), and the rate at which RNAs are incorrectly identified as a different RNA species (misidentification rate) (Fig. 5D) plotted as a function of the number of bits (*N*) in the binary words encoding RNA. In Figs. 5B and 5D, the upper dots are a simple binary code that includes all 2^N-1 possible binary words; the middle dots are the HD4 code where the Hamming distance separating words is 4; and the lower dots are the modified HD4 (MHD4) code where the number of '1' bits are kept at four. These are reversed in Fig. 5C.

The calling and misidentification rates are calculated with per bit error rates of 10% for the 1-->0 error and 4% for the 0-->1 error. Fig. 5E is as schematic diagram of the implementation of a MHD4 code for RNA identification. Each RNA species is first labeled with ~192 encoding probes that convert the RNA into a unique combination of readout sequences (Encoding hyb). These encoding probes each contain a central RNA targeting region flanked by two readout sequences, drawn from a pool of *N* different sequences, each associated with a specific hybridization round. Encoding probes for a specific RNA species contain a unique combination of four of the *N* readout sequences, which correspond to the four hybridization rounds where this RNA should read '1'. *N* subsequent rounds of hybridization with the fluorescent readout probes were used to probe the readout sequences (hyb 1, hyb 2, ..., hyb *N*). The bound probes were inactivated by photobleaching between successive rounds of hybridization. For clarity, only one possible pairing of the readout sequences is depicted here for the encoding probes; however, all possible pairs of the four readout sequences are used at the same frequency and distributed randomly along each cellular RNA in the actual experiments.

Fig. 11 shows a schematic description of a combinatorial labeling approach based on a simple binary code. In a conceptually simple labeling approach, 2^N-*1* different RNA

species can be uniquely encoded with all *N*-bit binary words (excluding the word with all '0's). In each hybridization round, FISH probes that are targeted to all RNA species that have a '1' in the corresponding bit are included. To increase the ability to discriminate RNA spots from background, each RNA is addressed with multiple FISH probes per hybridization round. Signal from the bound probes is extinguished before the next round of hybridization. This process continues for all *N* hybridization rounds (hyb 1, hyb 2, ...), and all 2^N-*1* RNA species can be identified by the unique on-off pattern of fluorescence signals in each hybridization round.

Fig. 12 shows schematic descriptions of Hamming distance and its use in the identification and correction of errors. Fig. 12A is a schematic representation of a Hamming distance of 4. Figs. 12B and 12C are schematics showing the ability of an encoding scheme with Hamming distance 4 to correct single-bit errors (Fig. 12B) or detect but not correct double-bit errors (Fig. 12C). Arrows highlight bits at which the indicated words differ. Two code words are separated by a Hamming distance of 4 if one of the words has to flip four bits from '1' to '0' or '0' to '1' to convert into the other word. Single-bit error correction is possible because if a measured word differs from a legitimate code word by only one bit, it is most likely an error that arises from misreading this code word, since the code words of all the other RNA species will differ from the measured word by at least three bits. In this case, the measured word can be corrected to a code word that differs by only one bit. If a measured word differs from a legitimate code word by two bits, this measured word can still be identified as an error, but correction is no longer possible since more than one legitimate code word differs from this measured word by two bits.

Fig. 13 shows production of the library of encoding probes. An array-synthesized complex oligopool, containing ~100k sequences, is used as a template for the enzymatic amplification of the encoding probes for different experiments. Each template sequence in the oligopool contains a central target region that can bind to a cellular RNA, two flanking readout sequences, and two flanking index primers. In the first step, the required template molecules for a specific experiment are selected and amplified with an indexed PCR reaction. To allow amplification via *in vitro* transcription, a T7 promotor is added to the PCR products during this step. In the second step, RNA is amplified from these template molecules via *in vitro* transcription. In the third step, this RNA is reverse transcribed back into DNA. In the final step, the template RNA is removed via alkaline hydrolysis, leaving only the desired ssDNA probes. This protocol produces ~2 nmol of complex pools of encoding probes

containing ~20,000 different sequences for the 140-gene experiments or ~100,000 different sequences for the 1001-gene experiments.

EXAMPLE 5

This example illustrates the measurement of 140 genes with MERFISH using a 16-bit MHD4 Code. To test the feasibility of this error-robust, multiplexed imaging approach, this example uses a 140-gene measurement on human fibroblast cells (IMR90) using a 16-bit MHD4 code to encode 130 RNA species while leaving 10 code words as misidentification controls (Fig. 20). After each round of hybridization with the fluorescent readout probes, cells were imaged by conventional wide-field imaging with an oblique-incidence illumination geometry. Fluorescent spots corresponding to individual RNAs were clearly detected and were then efficiently extinguished via a brief photobleaching step (Fig. 6A). The sample was stable throughout the 16 rounds of iterative labeling and imaging. The change in the number of fluorescent spots from round to round matched the predicted change based on the relative abundances of RNA species targeted in each round derived from bulk sequencing, and a systematic decreasing trend with increasing number of hybridization rounds was not observed (Fig. 14A). The average brightness of the spots varied from round to round with a standard deviation of 40%, likely due to different binding efficiencies of the readout probes to the different readout sequences on the encoding probes (Fig. 14B). Only a small, systematic decreasing trend in the spot brightness with increasing hybridization rounds was observed, which was on average 4% per round (Fig. 14B).

Next, binary words were constructed from the observed fluorescent spots based on their on-off patterns across the 16 hybridization rounds (Fig. 6B-6D). If the word exactly matched one of the 140 MHD4 code words (exact matches) or differed by only one bit (error-correctable matches), it was assigned to the corresponding RNA species (Fig. 6D). Within the single cell depicted in Figs. 6A and 6B, more than 1500 RNA molecules corresponding to 87% of the 130 encoded RNA species were detected after error correction (Fig. 6E). Similar observations were made in ~400 cells from 7 independent experiments. On average, ~4 times as many RNA molecules and ~2 times as many RNA species were detected per cell after error correction as compared with the values obtained before error correction (Fig. 15).

Two types of errors can occur in the copy number measurement of each RNA species: 1) Some molecules of this RNA species are not detected, leading to a drop in calling rate, and 2) some molecules from other RNA species are misidentified as this RNA species. To assess the extent of misidentification, the 10 misidentification control words were utilized, i.e., code words that were not associated with any cellular RNA. Although matches to these control

words were observed, they occurred far less frequently than the real RNA-encoding words: 95% of the 130 RNA-encoding words were counted more frequently than the median count for these control words. Moreover, it was typically found that the ratio of the number of exact matches to the number of matches with one-bit errors for a real RNA-encoding word was substantially higher than the same ratios observed for the misidentification controls, as expected (Figs. 16A and 16B). Using this ratio as a measure of the confidence in RNA identification, it was found that 91% of the 130 RNA species had a confidence ratio greater than the maximum confidence ratio observed for the misidentification controls (Fig. 6F), demonstrating a high accuracy of RNA identification. Subsequent analyses were conducted only on these 91% of genes.

To estimate the calling rate, the error-correction ability of the MHD4 code was utilized to determine the 1-->0 error rates (10% on average) and 0-->1 error rates (4% on average) for each hybridization round (Figs. 16C and 16D). Using these error rates, an ~80% calling rate for individual RNA species after error correction was estimated, i.e. ~80% of the fluorescent spots corresponding to a RNA species were decoded correctly (Fig. 16E). It is noted that although the remaining 20% of spots contributed to a loss in detection efficiency, most of them did not cause species misidentification because they were decoded as double-bit error words and discarded.

To test for potential technical bias in these measurements, the same 130 RNAs species were probed with a different MHD4 codebook by shuffling the code words among different RNA species (Fig. 20) and changing the encoding probe sequences. Measurements with this alternative code gave similar misidentification and calling rates (Fig. 17). The copy numbers of individual RNA species per cell measured with these two codebooks showed excellent agreement with a Pearson correlation coefficient of 0.94 (Fig. 6G), indicating that the choice of encoding scheme did not bias the measured counts.

In order to validate the copy numbers derived from the MERFISH experiments, conventional smFISH measurements were performed on 15 of the 130 genes, selected from the full measured abundance range of three orders of magnitude. For each of these genes, both the average copy number and the copy number distribution across many cells agreed quantitatively between the MERFISH and conventional smFISH measurements (Fig. 18A and 18B). The ratio of the copy numbers determined by these two approaches was 0.82 +/- 0.06 (mean +/- SEM across the 15 measured RNA species, Fig. 18B), which agreed with the estimated 80% calling rate for the multiplexed imaging approach. The quantitative match between this ratio and the estimated calling rate over the full measured abundance range

additionally supports the assessment that the misidentification error was low. Given that the agreement between the MERFISH and conventional smFISH results extended to the genes at the lowest measured abundance (<1 copy per cell, Fig. 18B), it was estimated that the measurement sensitivity was at least 1 copy per cell.

As a final validation, the abundance of each RNA species averaged over hundreds of cells was compared to those obtained from a bulk RNA sequencing measurement that were performed on the same cell line. The imaging results correlated remarkably well with bulk sequencing results with a Pearson correlation coefficient of 0.89 (Fig. 6H).

Fig. 6 shows simultaneous measurement of 140 RNA species in single cells using MERFISH with a 16-bit MHD4 code. Fig. 6A shows images of RNA molecules in an IMR90 cell after each hybridization round (hyb 1 – hyb 16). The image after photobleaching (bleach 1) demonstrated efficient removal of fluorescent signals between hybridizations. Fig. 6B shows the localizations of all detected single molecules in this cell colored based on their measured binary words. Inset: the composite fluorescent image of the 16 hybridization rounds for the boxed sub-region with numbered circles indicating potential RNA molecules. A circle indicates an unidentifiable molecule, the binary word of which does not match any of the 16-bit MHD4 code words even after error correction. Fig. 6C shows fluorescent images from each round of hybridization for the boxed sub-region in Fig. 6B with circles indicating potential RNA molecules. Fig. 6D shows corresponding words for the spots identified in Fig. 6C. Crosses represent the corrected bits. Fig. 6E shows the RNA copy number for each gene observed without (lower) or with (higher) error correction in this cell. Fig. 6F shows the confidence ratio measured for the 130 RNA species (left) and the 10 misidentification control words (right) normalized to the maximum value observed from the misidentification controls (dashed line). Fig. 6G is a scatter plot of the average copy number of each RNA species per cell measured with two shuffled codebooks of the MHD4 code. The Pearson correlation coefficient is 0.94 with a p-value of 1×10^{-53}. The dashed line corresponds to the $y = x$ line. Fig. 6H is a scatter plot of the average copy number of each RNA species per cell versus the abundance determined by bulk sequencing in fragments per kilobase per million reads (FPKM). The Pearson correlation coefficient between the logarithmic abundances of the two measurements was 0.89 with a p-value of 3×10^{-39}.

Fig. 14 shows the number and average brightness of the fluorescent spots detected in the 16 rounds of hybridization before and after photobleaching. Fig. 14A shows the number of fluorescent spots observed per cell before (higher) and after (lower) photobleaching as a function of hybridization round averaged across all measurements with the first 16-bit MHD4

code. Photobleaching reduces the number of fluorescent spots by two or more orders of magnitude. Hybridization rounds without lower bars represent rounds in which no molecules were observed after bleaching. Also depicted is the expected change in the number of fluorescent spots from round to round (circles) predicted based on the relative abundances of the RNA species targeted in each hybridization round derived from bulk RNA sequencing. The average discrepancy between the observed and predicted number of spots for each hybridization is only 15% of the mean number of spots. This discrepancy does not systematically increase with the number of hybridization rounds. Fig. 14B shows the average brightness of the identified fluorescent spots in each hybridization round averaged across all measurements with the first 16-bit MHD4 code both before (upper) and after (lower) photobleaching. Brightness varies by 40% (standard deviation) across different hybridization rounds. The variation pattern is reproducible between experiments with the same code, likely due to differences in the binding efficiency of the readout probes to the different readout sequences. There is a small systematic trend of decrease in the brightness with increasing hybridization rounds, which is on average 4% per round. Photobleaching extinguishes fluorescence to a level similar to that of the autofluorescence of the cell.

Fig. 15 shows error correction substantially increases the numbers of RNA molecules and RNA species detected in individual cells. Fig. 15A shows a histogram of the ratio of the total number of molecules detected per cell with error correction to the number measured without error correction. Fig. 15B is a histogram of the total number of RNA species detected in each cell with error correction to that without error correction. Both ratios are determined for ~200 cells and the histograms are constructed from these ratios.

Fig. 16 shows characterization of the misidentification and calling rates of RNA species for the 140-gene experiments using a specific 16-bit MHD4 code. Fig. 16A shows the number of measured words exactly matching the code word corresponding to FLNA, represented by the bar in the center of the circle, and the number of measured words with one-bit error compared to the code word of FLNA, represented by the 16 bars on the circle. Fig. 16B is the same as Fig. 16A, but for a code word that was not assigned to any RNA, i.e., a misidentification control word. The solid lines connect the exact match to one-bit error words that are generated by 1-->0 errors. Based on the observation that the ratio of the number of exact matches to the number of error-correctable matches for a real RNA-encoding word was typically substantially higher than the same ratios observed for the misidentification controls, this ratio was defined as a confidence ratio for RNA identification. The confidence ratio measured for all 130 RNA species (center bar) and 10 misidentification

control words not assigned to any RNA (outer bars) using this 16-bit MHD4 code is show in Fig. 6F. Fig. 16C and 16D show the average error rates for the 1-->0 error (Fig. 16C) and 0-->1 error (Fig. 16D) for each hybridization round. Fig. 16E shows the calling rate for each RNA species estimated from the 1-->0 and 0-->1 error rates. Genes are sorted from left to right based on the measured abundance, which spans three orders of magnitude. The calling rates are largely independent of the abundance of the gene.

Fig. 17 shows characterization of the misidentification and calling rates for a second 16-bit MHD4 code. In this second encoding scheme, the 140 code words were shuffled among different RNA species and changed the encoding probe sequences. Fig. 17A shows the normalized confidence ratio measured for the 130 RNA species (left) and the 10 misidentification control words not assigned to any RNA (right). The normalized confidence ratio is determined the same way as in Fig. 6F. Figs. 17B and 17C show the average error rates determined for the 1-->0 error (Fig. 17B) and 0-->1 error (Fig. 17C) for each hybridization round. Fig. 17D shows the calling rate determined for each RNA species estimated from the 1-->0 and 0-->1 error rates. Genes are sorted from left to right based on the measured abundance.

Fig. 18 shows a comparison of the MERFISH measurements with conventional smFISH results for a subset of genes. Fig. 18A shows the distributions of RNA copy numbers in single cells for three example genes KIAA1199, DYNC1H1, and LMTK2 in the high, medium, and low abundance ranges, respectively. Lighter bars: distributions constructed from ~400 cells in the 140-gene measurements using the MHD4 codes. Darker bars: distributions constructed from ~100 cells in the conventional smFISH measurements. Fig. 18B shows a comparison of the average RNA copy numbers per cell measured in the 140-gene experiments using the MHD4 codes to those determined by conventional smFISH for 15 genes. The average ratio of the copy number measured using the MHD4 measurements to that measured using conventional smFISH was 0.82 +/- 0.06 (mean +/- SEM across 15 genes). The dashed line corresponds to the $y = x$ line.

Fig. 20 shows two different codebooks for the 140-gene experiments. The specific code words of the 16-bit MHD4 code assigned to each RNA species studied in the two shuffles of the 140-gene experiment. The "Genes" columns contain the name of the gene. The "Codewords" columns contain the specific binary word assigned to each gene.

EXAMPLE 6

This example is generally directed to high-throughput analysis of cell-to-cell variation in gene expression. The MERFISH approach allows parallelization of measurements of many individual RNA species and co-variation analysis between different RNA species. In this example, the parallelization aspect was first illustrated by examining the cell-to-cell variation in the expression level of each of the measured genes (Fig. 7A). To quantify the measured variation, the Fano factors, defined as the ratio of the variance to the mean RNA copy number, were computed for all measured RNA species. The Fano factors substantially deviated from 1, the value expected for a simple Poisson process, for many genes and exhibited an increasing trend with the mean RNA abundance (Fig. 7B). This trend of increasing Fano factors with mean RNA abundance can be explained by changes in the transcription rate and/or promoter off-switching rates but not by changes in the promoter on-switching rate.

Moreover, several RNA species were identified with substantially larger Fano factors than this average trend. For example, it was found that SLC5A3, CENPF, MKI67, TNC and KIAA1199 displayed Fano factor values substantially higher than those of the other genes expressed at similar abundance levels. The high variability of some of these genes can be explained by their association with the cell cycle. For example, two of these particularly 'noisy' genes MKI67 and CENPF were both annotated as cell-cycle related genes, and based on their bimodal expression (Fig. 7C), it is proposed that their transcription is strongly regulated by the cell cycle. Other high-variability genes did not show the same bimodal expression patterns and are not known to be associated with the cell cycle.

Analysis of co-variations in the expression levels of different genes can reveal which genes are co-regulated and elucidate gene regulatory pathways. At the population level, such analysis often requires the application of external stimuli to drive gene expression variation; hence, correlated expression changes can be observed among genes that share common regulatory elements influenced by the stimuli. At the single-cell level, one can take advantage of the natural stochastic fluctuations in gene expression for such analysis and can thus study multiple regulatory networks without having to stimulate each of them individually. Such co-variation analysis can constrain regulatory networks, suggest new regulatory pathways, and predict function for unannotated genes based on associations with co-varying genes.

This approach was applied to the 140-gene measurements and the ~10,000 pairwise correlation coefficients describing how the expression levels of each pair of genes co-varied from cell to cell were examined. Many of the highly variable genes showed tightly correlated

or anti-correlated variations (Fig. 7C). To better understand the correlations for all gene pairs, a hierarchical clustering approach was adopted to organize these genes based on their correlation coefficients (Fig. 7D). From the cluster tree structure, seven groups of genes with substantially correlated expression patterns were identified (Fig. 7D). Within each of the seven groups, every gene showed significantly stronger average correlation with other members of the group than with genes outside the group. To further validate and understand these groups, gene ontology (GO) terms enriched in each of these seven groups were identified. Notably, the enriched GO terms within each group shared similar functions and were largely unique to each group (Fig. 7E), validating the notion that the observed co-variation in expression reflects some commonalities in the regulation of these genes.

This example describes two of these groups as illustrative examples. The predominant GO terms associated with Group 1 were terms associated with the extracellular matrix (ECM) (Figs. 7D and 7E). Notable members of this group included ECM components, such as FBN1, FBN2, COL5A, COL7A and TNC, and glycoproteins linking the ECM and cell membranes, such as VCAN and THBS1. The group also included an unannotated gene, KIAA1199, which may predict to play a role in ECM metabolism based on its association with this cluster. Indeed, this gene has recently been identified as an enzyme involved in the regulation of hyaluronan, a major sugar component of the ECM.

Group 6 contained many genes that encode vesicle transport proteins and proteins associated with cell motility (Figs. 7D and 7E). The vesicle transport genes included microtubule motors and related genes DYNC1H, CKAP1, and factors associated with vesicle formation and trafficking, like DNAJC13 and RAB3B. Again, an unannotated gene, KIAA1462, was found within this cluster. Based on its strong correlation with DYNC1H1 and DNAJC13, it is predicted that this gene may be involved in vesicle transport. The cell motility genes in this group included actin-binding proteins like AFAP1, SPTAN1, SPTBN1, and MYH10, and genes involved in the formation of adhesion complexes, like FLNA and FLNC. Several GTPase-associated factors involved in the regulation of cell motility, attachment and contraction also fell into this group, including DOCK7, ROCK2, IQGAP1, PRKCA, and AMOTL1. The observation that some cell motility genes correlated with vesicle transport genes is consistent with the role of vesicle transport in cell migration. An additional interesting feature of group 6 is that a subset of these genes, in particular those related to cell motility, were anti-correlated with members of the ECM group discussed above (Fig. 7D). This anti-correlation may reflect regulatory interactions that mediate switching of cells between adherent and migratory states.

Fig. 7 shows cell-to-cell variations and pairwise correlations for the RNA species determined from the 140 gene measurements. Fig. 7A shows a comparison of gene expression levels in two individual cells. Fig. 7B shows Fano factors for individual genes. Error bars represent standard error of the mean determined from 7 independent data sets. Fig. 7C shows Z-scores of the expression variations of four example pairs of genes showing correlated (top two) or anti-correlated (bottom two) variation for 100 randomly selected cells. The Z-score is defined as the difference from the mean normalized by the standard deviation. Fig. 7D is a matrix of the pairwise correlation coefficients of the cell-to-cell variation in expression for the measured genes, shown together with the hierarchical clustering tree. The seven groups identified by a specific threshold on the cluster tree (dashed line) are indicated by the black boxes in the matrix and lines on the tree, with grey lines on the tree indicating ungrouped genes. Different threshold choices on the cluster tree could be made to select either smaller subgroups with tighter correlations or larger super-groups containing more weakly coupled subgroups. Two of the seven groups are enlarged on the right. Fig. 7E shows enrichment of 30 selected, statistically significantly enriched GO terms in the seven groups. Enrichment refers to the ratio of the fraction of genes within a group that have the specific GO term to the fraction of all measured genes having that term. Not all of the GO terms presented here are in the top 10 list.

EXAMPLE 7

This example illustrates mapping spatial distributions of RNAs. As an imaging based approach, MERFISH also allowed the investigation of the spatial distributions of many RNA species simultaneously. Several patterns emerged from the visual inspection of individual genes, with some RNA transcripts enriched in the perinuclear region, some enriched in the cell periphery, and some scattered throughout the cell (Fig. 8A). To identify genes with similar spatial distributions, the correlation coefficients for the spatial density profiles of all pairs of RNA species were determined, and these RNAs were organized based on the pairwise correlations again using the hierarchical clustering approach. The correlation coefficient matrix showed groups of genes with correlated spatial organizations, and the two most notable groups with the strongest correlations are indicated in Fig. 8B. Group I RNAs appeared enriched in the perinuclear region whereas group II RNAs appeared enriched near the cell peripheral region (Fig. 8C). Quantitative analysis of the distances between each RNA molecule and the cell nucleus or the cell periphery indeed confirmed this visual impression (Fig. 8D).

Group I contained genes encoding extracellular proteins such as FBN1, FBN2 and THSB1, secreted proteins such as PAPPA, and integral membrane proteins such as LRP1 and GPR107. These proteins have no obvious commonalities in function. Rather a GO analysis showed significant enrichment for location terms, such as extracellular region, basement membrane, or perivitelline space (Fig. 8E). To reach these locations, proteins must pass through the secretion pathway, which often requires translation of mRNA at the endoplasmic reticulum (ER). Thus, it is proposed that the spatial pattern that were observed for these mRNAs reflects their co-translational enrichment at the ER. The enrichment of these mRNAs in the perinuclear region (Figs. 8C and 8D, lighter shading), where the rough ER resides, supports this conclusion.

Group II contained genes encoding the actin-binding proteins, including filamins FLNA and FLNC, talin TLN1, and spectrins SPTAN1 and SPTBN1; the microtubule-binding protein CKAP5; and the motor proteins MYH10 and DYNC1H1. This group was enriched with GO terms such as cortical actin cytoskeleton, actin filament binding, and cell-cell adherens junction (Fig. 8E). Beta-actin mRNA may be enriched near the cell periphery in fibroblasts as are mRNAs that encode members of the actin-binding Arp2/3 complex. The enrichment of group II mRNAs in the peripheral region of the cells (Figs. 8C and 8D) suggests that the spatial distribution of the Group II genes might be related to the distribution of actin cytoskeleton mRNAs.

Fig. 8 shows distinct spatial distributions of RNAs observed in the 140-gene measurements. Fig. 8A shows examples of the spatial distributions observed for four different RNA species in a cell. Fig. 8B is a matrix of the pairwise correlation coefficients describing the degree with which the spatial distributions of each gene pair is correlated, shown together with the hierarchical clustering tree. Two strongly correlating groups are indicated by the black boxes on the matrix and shading on the tree. Fig. 8C shows the spatial distributions of all RNAs in the two groups in two example cells. Lighter symbols: group I genes; darker symbols: group II genes. Fig. 8D shows average distances for genes in group I and genes in group II to the cell edge or the nucleus normalized to the average distances for all genes. Error bars represent SEM across 7 data sets. Fig. 8E shows enrichment of GO terms in each of the two groups.

EXAMPLE 8

This example illustrates measuring 1001 genes with a 14-bit MHD2 code. This example further increases the throughput of MERFISH measurements by simultaneously imaging ~1000 RNA species. This increase could be achieved with the MHD4 code by

increasing the number of bits per code word to 32 while maintaining the number of '1' bits per word at four (Fig. 5B). While the stability of the samples across many hybridization rounds (Fig. 14) suggests that such an extension is potentially feasible, an alternative approach is shown here that did not require an increase in the number of hybridizations by relaxing the error correction requirement but keeping the error detection capability. For example, by reducing the Hamming distance from 4 to 2, all 14-bit words could be used that contain four '1' bits to encode 1001 genes and these RNAs were probed with only 14 rounds of hybridization. However, because a single error can produce a word equally close to two different code words, error correction is no longer possible for this modified Hamming-distance-2 (MHD2) code. Hence, it was expected that the calling rate would be lower and the misidentification rate to be higher with this encoding scheme.

To evaluate the performance of this 14-bit MHD2 code, 16 of the 1001 possible code words were set aside as misidentification controls and used the remaining 985 words to encode cellular RNAs. Among these 985 RNAs included 107 RNA species probed in the 140-gene experiments as an additional control. The 1001-gene experiments were performed in IMR90 cells using a similar procedure as described above. To allow all encoding probes to be synthesized from a single 100,000-member oligopool, the number of encoding probes per RNA species was reduced to ~94. Fluorescent spots corresponding to individual RNA molecules were again detected in each round of hybridization with the readout probes and, based on their on-off patterns, these spots were decoded into RNA (Figs. 9A, 19A and 19B). 430 RNA species were detected in the cell shown in Fig. 9A, and similar results were obtained in ~200 imaged cells in 3 independent experiments.

As expected, the misidentification rate of this scheme was higher than that of the MHD4 code. 77% of all real RNA words were detected more frequently than the median count for the misidentification controls instead of the 95% value observed in the MHD4 measurements. Using the same confidence ratio analysis as described above, it was found that 73% (instead of 91% for the MHD4 measurements) of the 985 RNA species were measured with a confidence ratio larger than the maximum value observed for the misidentification controls (Fig. 19C). RNA copy numbers measured from these 73% RNA species showed excellent correlation with the bulk RNA sequencing results (Pearson correlation coefficient $r = 0.76$; Fig. 9B, black). It is worth noting that the remaining 27% of the genes still exhibit good, albeit lower, correlation with the bulk RNA sequencing data ($r = 0.65$; Fig. 9B, red), but the conservative measure of excluding them from further analysis was taken.

The lack of an error correction capability also decreased the calling rate of each RNA species: When comparing the 107 RNA species common in both the 1001-gene and 140-gene measurements, it was found that the copy numbers per cell of these RNA species were lower in the 1001-gene measurements (Figs. 9C and 19D). The total count of these RNAs per cell was ~1/3 of that observed in the 140-gene measurements. Thus the lack of error correction in the MHD2 code produced a ~3-fold decrease in the calling rate, which is consistent with the ~4-fold decrease in calling rate observed for the MHD4 code when error correction was not applied. As expected from the quantitative agreement between 140-gene measurements and conventional smFISH results, comparison of the 1001-gene measurements with conventional smFISH results for 10 RNA species also indicated a ~3-fold drop in calling rate (Fig. 18C). Despite the expected reduction in calling rate, the good correlations found between the copy numbers observed in the 1001-gene measurements and those observed in the 140-gene measurements, as well as in conventional smFISH and bulk RNA sequencing measurements, indicates that the relative abundance of these RNAs can be quantified with the MHD2 encoding scheme.

Simultaneously imaging ~1000 genes in individual cells substantially expanded the ability to detect co-regulated genes. Fig. 10A shows the matrix of pairwise correlation coefficients determined from the cell-to-cell variations in the expression levels of these genes. Using the same hierarchical clustering analysis as described above, ~100 groups of genes with correlated expression were identified. Remarkably, nearly all of these ~100 groups showed statistically significant enrichment of functionally related GO terms (Fig. 10B). These included some of the groups identified in the 140-gene measurements, such as the group associated with cell replication genes and the group associated with cell motility genes (Figs. 10A and 10B, groups 7 and 102), as well as many new groups. The groups identified here included 46 RNA species lacking any previous GO annotations, for which function based on their group association may be hypothesized. For example, KIAA1462 is part of the cell motility group, as also shown in the 140-gene experiments, suggesting a potential role of this gene in cell motility (Fig. 10A, group 102). Likewise, KIAA0355 is part of a new group enriched in genes associated with heart development (Fig. 10A, group 79), and C17orf70 is part of a group associated with ribosomal RNA processing (Fig. 10A, group 22). Using these groupings, cellular functions for 61 transcription factors and other partially annotated proteins of unknown functions may be hypothesized. For example, the transcription factors Z3CH13 and CHD8 are both members of the cell motility group, suggesting their potential role in the transcriptional regulation of cell motility genes.

Fig. 9 shows simultaneous measurements of 1001 genes in single cells using MERFISH with a 14-bit MHD2 code. Fig. 9A shows the localizations of all detected single molecules in a cell colored based on their measured binary words. Inset: the composite, false-colored fluorescent image of the 14 hybridization rounds for the boxed sub-region with numbered circles indicating potential RNA molecules. Circles indicate unidentifiable molecules, the binary words of which do not match any of the 14-bit MHD2 code words. Images of individual hybridization round are shown in Fig. 19A. Fig. 9B is a scatter plot of the average copy number per cell measured in the 1001-gene experiments versus the abundance measured via bulk sequencing. The upper symbols are for the 73% of genes detected with confidence ratios higher than the maximum ratio observed for the misidentification controls. The Pearson correlation coefficient is 0.76 with a p-value of 3×10^{-133}. The lower symbols are for the remaining 27% of genes. The Pearson correlation coefficient is 0.65 with a p-value of 3×10^{-33}. Fig. 9C is a scatter plot of the average copy number for the 107 genes shared in both the 1001-gene measurement with the MHD2 code and the 140-gene measurement with the MHD4 code. The Pearson correlation coefficient is 0.89 with a p-value of 9×10^{-30}. The dashed line corresponds to the $y = x$ line.

Fig. 10 shows co-variation analysis of the RNA species measured in the 1001-gene measurements. Fig. 10A is a matrix of all pairwise correlation coefficients of the cell-to-cell variation in expression for the measured genes shown with the hierarchical clustering tree. The ~100 identified groups of correlated genes are indicated by shading on the tree. Zoom in of four of the groups described in the text are shown on the right. Fig. 10B is an enrichment of 20 selected, statistically significantly enriched GO terms in the four groups.

Fig. 19 shows decoding and error assessment of the 1001-gene experiments. Fig. 19A shows images of the boxed sub-region of the cell in Fig. 9A for each of 14 hybridization rounds. The final panel is a composite image of these 14 rounds. Circles indicate fluorescent spots that have been identified as potential RNA molecules. Some circles in the composite image indicate unidentifiable molecules, the binary words of which do not match any of the 14-bit MHD2 code words. Fig. 9B shows the corresponding binary word for each of the spots identified in Fig. 9A with the RNA species to which it is decoded. 'unidentified' implies that the measured binary word does not match any of the 1001 code words. Fig. 19C shows the normalized confidence ratios measured for the 985 RNA species (left) and the 16 misidentification control words not targeted to any RNA (right). The normalized confidence ratio is defined as in Fig. 6F. Fig. 19D shows a histogram of the reduction in detected abundance of 107 genes present in both the 1001-gene experiments and the 140-gene

experiments. "Fold decrease in copy number" is defined as the average number of RNA molecules per cell for each species measured in the 140-gene experiments divided by the corresponding average number measured in the 1001-gene experiments.

Fig. 18C is a comparison of the average RNA copy numbers per cell measured in the 1001-gene experiments using the MHD2 code to those determined by conventional smFISH for 10 genes. The average ratio of the copy number measured using the MHD2 measurements to that measured using conventional smFISH was 0.30 +/- 0.05 (mean +/- SEM across 10 genes). The dashed line corresponds to the $y = x$ line and the dotted line corresponds to the $y = 0.30x$ line.

EXAMPLE 9

The above examples illustrate a highly multiplexed detection scheme for systems-level RNA imaging in single cells. Using combinatorial labeling, sequential hybridization and imaging, and two different error-robust encoding schemes, either 140 or 1001 genes in hundreds of individual human fibroblast cells were simultaneously imaged. Of the two encoding schemes presented here, the MHD4 code is capable of both error detection and error correction, and hence can provide a higher calling rate and a lower misidentification rate than the MHD2 code, which instead can only detect but cannot correct errors. MHD2, on the other hand, provides a faster scaling of the degree of multiplexing with the number of bits than MHD4. Other error-robust encoding schemes can also be used for such multiplexed imaging, and experimenters can set the balance between detection accuracy and ease of multiplexing based on the specific requirements of the experiments.

By increasing the number of bits in the code words, it should be possible to further increase the number of detectable RNA species using MERFISH with, for example, a MHD4 or MHD2 code. For example, using the MHD4 code with 32 total bits and four or six '1' bits would increase the number of addressable RNA species to 1,240 or 27,776, respectively. The latter is the approximate scale of the human transcriptome. The predicted misidentification and calling rates are still reasonable for the 32-bit MHD4 code (shown in Figs. 5C and 5D for the MHD4 code with four '1' bits and similar rates were calculated for the MHD4 code with six '1' bits). If more accurate measurements are desired, an additional increase in the number of bits would allow the use of encoding schemes with a Hamming distance greater than 4, further enhancing the error detection and correction capability. While an increase in the number of bits by adding more hybridization rounds would increase the data collection time and potentially lead to sample degradation, these problems could be mitigated by utilizing multiple colors to readout multiple bits in each round of hybridization.

As the degree of multiplexing is increased, it is important to consider the potential increase in the density of RNAs that need to be resolved in each round of imaging. Based on the imaging and sequencing results, it can be estimated that including the whole transcriptome of the IMR90 cells would lead to a total RNA density of ~200 molecules/micrometer^3. Using the current imaging and analysis methods, 2-3 molecules/ micrometer^3 per hybridization round could be resolved, which would reach a total RNA density of ~20 molecules/micrometer^3 after 32 rounds of hybridization. This density should allow all but the top 10% most expressed genes to be imaged simultaneously or a subset of genes with even higher expression levels to be included. By utilizing more advanced image analysis algorithms to better resolve overlapping images of individual molecules, such as compressed sensing, it is possible to extend the resolvable density by ~4-fold and thus allow all but the top 2% most expressed genes to be imaged all together.

These examples have illustrated the utility of the data derived from highly multiplexed RNA imaging by using co-variation and correlation analysis to reveal distinct sub-cellular distribution patterns of RNAs, to constrain gene regulatory networks, and to predict functions for many previously unannotated or partially annotated genes with unknown functions. Given its ability to quantify RNAs across a wide range of abundances without amplification bias while preserving native context, systems and methods such as MERFISH will allow many applications of *in situ* transcriptomic analyses of individual cells in culture or complex tissues.

EXAMPLE 10

Following are various materials and methods used in the above examples.

Probe design. Each RNA species in the target set was randomly assigned a binary code word either from all 140 possible code words of the 16-bit MHD4 code or from all 1001 possible code words of the 14-bit MHD2 code.

Array-synthesized oligopools were used as templates to make the encoding probes. The template molecule for each encoding probe contained three components: i) a central targeting sequence for *in situ* hybridization to the target RNA, ii) two flanking readout sequences designed to hybridize each of two distinct readout probes, and iii) two flanking primer sequences to allow enzymatic amplification of the probes (Fig. 13). The readout sequences were taken from the 16 possible readout sequences, each corresponding to one hybridization round. The readout sequences were assigned to the encoding probes such that for any RNA species each of the 4 readout sequences were distributed uniformly along the length of the target RNA and appeared at the same frequency. Template molecules for the

140-gene library also included a common 20-nucleotide (nt) priming region between the first PCR primer and the first readout sequence. This priming sequence was used for the reverse transcription step described below.

Multiple experiments were embedded in a single array-synthesized oligopool, and PCR was used to selectively amplify only the oligos required for a specific experiment. Primer sequences for this indexed PCR reaction were generated from a set of orthogonal 25-nt sequences. These sequences were trimmed to 20 nt and selected for i) a narrow melting temperature range (70 °C to 80 °C), ii) the absence of consecutive repeats of 3 or more identical nucleotides, and iii) the presence of a GC clamp, i.e. one of the two 3' terminal bases must be G or C. To further improve specificity, these sequences were then screened against the human transcriptome using BLAST+, and primers with 14 or more contiguous bases of homology were eliminated. Finally, BLAST+ was again used to identify and exclude primers that had an 11-nt homology region at the 3' end of any other primer or a 5-nt homology region at the 3' end of the T7 promoter. The forward primer sequences (Primer 1) were determined as described above, whereas the reverse primers each contain a 20-nt sequence as described above plus a 20-nt T7 promoter sequence to facilitate amplification via *in vitro* transcription (Primer 2). The primer sequences used in the 140-gene and 1001-gene experiments are listed below.

Table 2

Experiment Name	Primer 1 Sequence (Index Primer 1)	Primer 2 Sequence (T7 promoter plus the reverse complement of Index Primer 2)
140-gene Codebook 1	GTTGGTCGGCACTTGGGTGC (SEQ ID NO: 18)	TAATACGACTCACTATAGGGAAAGCCGGTTCATCCGGTGG (SEQ ID NO: 21)
140-gene Codebook 2	CGATGCGCCAATTCCGGTTC (SEQ ID NO: 19)	TAATACGACTCACTATAGGGTGATCATCGCTCGCGGGTTG (SEQ ID NO: 22)
1001-gene	CGCGGGCTATATGCGAACCG (SEQ ID NO: 20)	TAATACGACTCACTATAGGGCGTGGAGGGCATACAACGC (SEQ ID NO: 23)

30-nt-long readout sequences were created by concatenating fragments of the same orthogonal primer set generated above by combining one 20-nt primer with a 10-nt fragment of another. These readout sequences were then screened, using BLAST+, for orthogonality with the index primer sequences and other readout sequences (no more than 11 nt of

homology) and for potential off-target binding sites in the human genome (no more than 14 nt of homology). Fluorescently labeled readout probes with sequences complementary to the readout sequences were used to probe these readout sequences, one in each hybridization round. All used readout probes sequences are listed below.

Table 3

Bit	Readout probes	
1	CGCAACGCTTGGGACGGTTCCAATCGGATC/3Cy5Sp/	SEQ ID NO: 24
2	CGAATGCTCTGGCCTCGAACGAACGATAGC/3Cy5Sp/	SEQ ID NO: 25
3	ACAAATCCGACCAGATCGGACGATCATGGG/3Cy5Sp/	SEQ ID NO: 26
4	CAAGTATGCAGCGCGATTGACCGTCTCGTT/3Cy5Sp/	SEQ ID NO: 27
5	GCGGGAAGCACGTGGATTAGGGCATCGACC/3Cy5Sp/	SEQ ID NO: 28
6	AAGTCGTACGCCGATGCGCAGCAATTCACT/3Cy5Sp/	SEQ ID NO: 29
7	CGAAACATCGGCCACGGTCCCGTTGAACTT/3Cy5Sp/	SEQ ID NO: 30
8	ACGAATCCACCGTCCAGCGCGTCAAACAGA/3Cy5Sp/	SEQ ID NO: 31
9	CGCGAAATCCCCGTAACGAGCGTCCCTTGC/3Cy5Sp/	SEQ ID NO: 32
10	GCATGAGTTGCCTGGCGTTGCGACGACTAA/3Cy5Sp/	SEQ ID NO: 33
11	CCGTCGTCTCCGGTCCACCGTTGCGCTTAC/3Cy5Sp/	SEQ ID NO: 34
12	GGCCAATGGCCCAGGTCCGTCACGCAATTT/3Cy5Sp/	SEQ ID NO: 35
13	TTGATCGAATCGGAGCGTAGCGGAATCTGC/3Cy5Sp/	SEQ ID NO: 36
14	CGCGCGGATCCGCTTGTCGGGAACGGATAC/3Cy5Sp/	SEQ ID NO: 37
15	GCCTCGATTACGACGGATGTAATTCGGCCG/3Cy5Sp/	SEQ ID NO: 38
16	GCCCGTATTCCCGCTTGCGAGTAGGGCAAT/3Cy5Sp/	SEQ ID NO: 39

The readout probes used for the 140-gene libraries were probes 1 through 16. The readout probes used for the 1001-gene experiment were probes 1 through 14. /3Cy5Sp/ indicates a 3' Cy5 modification.

To design the central targeting sequences of the encoding probes, the abundance of different transcripts in IMR90 cells using Cufflinks v2.1, total RNA data from the ENCODE project, and human genome annotations from Gencode v18 were complied. Probes were designed from gene models corresponding to the most abundant isoform using OligoArray2.1 with the following constraints: the target sequence region is 30-nt long; the melting temperatures of the hybridized region of the probe and cellular RNA target is greater than 70 ^{o}C; there is no cross hybridization targets with melting temperatures greater than 72 ^{o}C; there is no predicted internal secondary structures with melting temperatures greater than 76 ^{o}C; and there is no contiguous repeats of 6 or more identical nucleotides. Melting temperatures were adjusted to optimize the specificity of these probes and minimize secondary structure while still producing sufficient numbers of probes for the libraries. To decrease computational cost, isoforms were divided into 1-kb regions for probe design. Using

BLAST+, all potential probes that mapped to more than one cellular RNA species were rejected. Probes with multiple targets on the same RNA were kept.

For each gene in the 140-gene experiments, 198 putative encoding probe sequences were generated by concatenating the appropriate index primers, readout sequences, and targeting regions as shown in Fig. 13. To address the possibility that concatenation of these sequences introduced new regions of homology to off-target RNAs, BLAST+ was used to screen these putative sequences against all human rRNA and tRNA sequences as well as highly expressed genes (genes with FPKM > 10,000). Probes with greater than 14 nt of homology to rRNAs or tRNAs or greater than 17 nt of homology to highly expressed genes were removed. After these cuts, there were ~192 (with a standard deviation of 2) probes per gene for both MHD4 codebooks used in the 140-gene experiments. The same protocol for the 1001-gene experiments was used, as follows: Starting with 96 putative targeting sequences per gene, ~94 (with a standard deviation of 6) encoding probes per gene were obtained after these additional homology cuts. The number of encoding probes per RNA was decreased for the 1001-gene experiments so that these probes could be synthesized from a single 100,000-member oligopool as opposed to two separate pools. Each encoding probe was designed to contain two of the four readout sequences associated with each code word, hence only half of the bound encoding probes can bind readout probe during any given hybridization round. ~192 or ~94 encoding probes per RNA were used to obtain high signal-to-background ratios for individual RNA molecules. The number of encoding probes per RNA could be substantially reduced but still allow single RNA molecules to be identified. In addition, increasing the number of readout sequences per encoding probe or using optical sectioning methods to reduce the fluorescence background may allow further reduction in the number of the encoding probes per RNA.

Two types of misidentification controls were designed. The first control (blank words) were not represented with encoding probes. The second type of control (no-target words) had encoding probes that were not targeting any cellular RNA. The targeting regions of these probes were composed of random nucleotide sequences subject to the same constraints used to design the RNA targeting sequences described above. Moreover, these random sequences were screened against the human transcriptome to ensure that they contain no significant homology (>14-nt) to any human RNA. The 140-gene measurements contained 5 blank words and 5 no-target words. The 1001-gene measurements contained 11 blank words and 5 no-target words.

Probe synthesis. The encoding probes were synthesized using the following steps, and this synthesis protocol is illustrated in Fig. 13.

Step 1: The template oligopool (CustomArray) was amplified via limited-cycle PCR on a Bio-Rad CFX96 using primer sequences specific to the desired probe set. To facilitate subsequent amplification via *in vitro* transcription, the reverse primer contained the T7 promoter. All primers were synthesized by IDT. This reaction was column purified (Zymo DNA Clean and Concentrator; D4003).

Step 2: The purified PCR products were then further amplified ~200-fold and converted into RNA via a high yield *in vitro* transcription according to the manufacturer's instructions (New England Biolabs, E2040S). Each 20 microliter reaction contained ~1 microgram of template DNA from above, 10 mM of each NTP, 1x reaction buffer, 1x RNase inhibitor (Promega RNasin, N2611) and 2 microliters of the T7 polymerase. This reaction was incubated at 37 ^{o}C for 4 hours to maximize yield. This reaction was not purified before the following steps.

Step 3: The RNA products from the above *in vitro* transcription reaction were then converted back into DNA via a reverse transcription reaction. Each 50 microliter reaction contained the unpurified RNA produce from Step 2 supplemented with 1.6 mM of each dNTP, 2 nmol of a reverse transcription primer, 300 units of Maxima H- reverse transcriptase (Thermo Scientific, EP0751), 60 units of RNasin, and a final 1x concentration of the Maxima RT buffer. This reaction was incubated at 50 ^{o}C for 45 minutes, and the reverse transcriptase was inactivated at 85 ^{o}C for 5 minutes. The templates for the 140-gene libraries contain a common priming region for this reverse transcription step; thus, a single primer was used for this step when creating these probes. Its sequence was CGGGTTTAGCGCCGGAAATG (SEQ ID NO: 40). A common priming region was not included for the 1001-gene library; thus, the reverse transcription was conducted with the forward primer: CGCGGGCTATATGCGAACCG (SEQ ID NO: 20).

Step 4: To remove the template RNA, 20 microliters of 0.25 M EDTA and 0.5 N NaOH was added to the above reaction to selectively hydrolyze RNA, and the sample was incubated at 95 ^{o}C for 10 minutes. This reaction was then immediately purified by column purification using a 100-microgram-capacity column (Zymo Research, D4030) and the Zymo Oligo Clean and Concentrator protocol. The final probes were eluted in 100 microliters of RNase-free deionized water, evaporated in a vacuum concentrator, and then resuspended in 10 μL of encoding hybridization buffer (see below). Probes were stored at -20 ^{o}C. Denaturing poly-acrylamide gel electrophoresis and absorption spectroscopy were used to

confirm the quality of the probes and revealed that this probe synthesis protocol converts 90-100% of the reverse-transcription primer into full length probe and of the probe that is constructed, 70-80% is recovered during the purification step.

Fluorescently labeled readout probes have sequences complementary to the readout sequences described above and a Cy5 dye attached at the 3' end. These probes were obtained from IDT and HPLC purified.

Sample preparation and labeling with encoding probes. Human primary fibroblasts (American Type Culture Collection, IMR90) were used in this work. These cells are relatively large and flat, facilitating wide-field imaging without the need for optical sectioning. Cells were cultured with Eagle's Minimum Essential Medium. Cells were plated on 22-mm, #1.5 coverslips (Bioptechs, 0420-0323-2) at 350,000 cells/coverslip and incubated at 37 °C with 5% CO_2 for 48-96 hours within petri dishes. Cells were fixed for 20 minutes in 4% paraformaldehyde (Electron Microscopy Sciences, 15714) in 1x phosphate buffered saline (PBS; Ambion, AM9625) at room temperature, reduced for 5 minutes with 0.1% w/v sodium borohydride (Sigma, 480886) in water to reduce background fluorescence, washed three times with ice-cold 1x PBS, permeabilized for 2 minutes with 0.5% v/v Triton (Sigma, T8787) in 1x PBS at room temperature, and washed three times with ice cold 1x PBS.

Cells were incubated for 5 minutes in encoding wash buffer comprising 2x saline-sodium citrate buffer (SSC) (Ambion, AM9763), 30% v/v formamide (Ambion, AM9342), and 2 mM vanadyl ribonucleoside complex (NEB, S1402S). 10 microliters of 100 micromolar(140-gene experiments) or 200 micromolar (1001-gene experiments) encoding probes in encoding hybridization buffer was added to the cell-containing coverslip and spread uniformly by placing another coverslip on top of the sample. Samples were then incubated in a humid chamber inside a 37 °C-hybridization oven for 18-36 hours. Encoding hybridization buffer is composed of encoding wash buffer supplemented with 1 mg/mL yeast tRNA (Life Technologies, 15401-011) and 10% w/v dextran sulfate (Sigma, D8906-50G).

Cells were then washed with primary encoding wash buffer, incubated at 47 °C for 10 minutes, and this wash was repeated for a total of three times. A 1:1000 dilution of 0.2-micrometer-diameter carboxylate-modified orange fluorescent beads (Life Technologies, F-8809) in 2xSSC was sonicated for 3 minutes and then incubated with the sample for 5 minutes. The beads were used as fiducial markers to align images obtained from multiple successive rounds of hybridization, as described below. The sample was washed once with 2xSSC, and then post-fixed with 4% v/v paraformaldehyde in 2xSSC at room temperature for 30 minutes. The sample was then washed three times with 2xSSC and either imaged

immediately or stored for no longer than 12 hours at 4 °C prior to imaging. All solutions were prepared as RNase-free.

MERFISH imaging with multiple successive rounds of hybridization. The sample coverslip was assembled into a Bioptech's FCS2 flow chamber, and the flow through this chamber was controlled via a home-built fluidics system composed of three computer-controlled 8-way valves (Hamilton, MVP and HVXM 8-5) and a computer-controlled peristaltic pump (Rainin, Dynamax RP-1). The sample was imaged on a home-built microscope constructed around an Olympus IX-71 body and a 1.45 NA, 100x oil immersion objective and configured for oblique incidence excitation. The objective was heated to 37 °C with a Bioptechs objective heater. Constant focus was maintained throughout the imaging process with a home-built, auto-focusing system. Illumination was provided at 641 nm, 561 nm, and 405 nm using solid state lasers (MPB communications, VFL-P500-642; Coherent, 561-200CWCDRH; and Coherent, 1069413/AT) for excitation of the Cy5-labeled readout probes, the fiducial beads, and nuclear counterstains, respectively. These lines were combined with a custom dichroic (Chroma, zy405/488/561/647/752RP-UF1) and the emission was filtered with a custom dichroic (Chroma, ZET405/488/561/647-656/752m). Fluorescence was separated with a QuadView (Photometrics) using the dichroics T560lpxr, T650lpxr, 750dcxxr (Chroma) and the emission filters ET525/50m, WT59550m-2f, ET700/75m, HQ770lp (Chroma) and imaged with an EMCCD camera (Andor, iXon-897). The camera was configured so that a pixel corresponds to 167 nm in the sample plane. The entire system was fully automated, so that imaging and fluid handling were performed for the entire experiment without user intervention.

Sequential hybridization, imaging, and bleaching proceeded as follows. 1 mL of 10 nM of the appropriate fluorescently labeled readout probe in readout hybridization buffer (2xSSC; 10% v/v formamide; 10% w/v dextran sulfate, and 2 mM vanadyl ribonucleoside complex) was flown across the sample, flow was stopped, and the sample was incubated for 15 minutes. Then 2 mL of readout wash buffer (2xSSC, 20% v/v formamide; and 2 mM vanadyl ribonucleoside complex) was flown across the sample, flow was stopped, and the sample was incubated for 3 minutes. 2 mL of imaging buffer comprising 2xSSC, 50 mM TrisHCl pH 8, 10% w/v glucose, 2 mM Trolox (Sigma-Aldrich, 238813), 0.5 mg/mL glucose oxidase (Sigma-Aldrich, G2133), and 40 microgram/mL catalase (Sigma-Aldrich, C30) was flown across the sample. Flow was then stopped, and then approximately 75 to 100 regions were exposed to ~25 mW 642-nm and 1 mW of 561-nm light and imaged. Each region was 40 micrometers by 40 micrometers. The laser powers were measured at the microscope

backport. Because the imaging buffer is sensitive to oxygen, the ~50 mL of imaging buffer used for a single experiment was made fresh at the beginning of the experiment and then stored under a layer of mineral oil throughout the measurement. Buffer stored in this fashion was stable for more than 24 hours.

After imaging, the fluorescence of the readout probes was extinguished via photobleaching. The sample was washed with 2 mL of photobleaching buffer (2xSSC and 2 mM vanadyl ribonucleoside complex), and each imaged region of the sample was exposed to 200 mW of 641-nm light for 3 s. To confirm the efficacy of this photobleaching treatment, imaging buffer was reintroduced, and the sample was imaged as described above.

The above hybridization, imaging, and photobleaching process was repeated either 16 times for the 140-gene measurements using the MHD4 code or 14 times for the 1001-gene measurements using the MHD2 code. An entire experiment was typically completed in ~20 hours.

Following completion of imaging, 2 mL of a 1:1000 dilution of Hoescht (ENZ-52401) in 2xSSC was flown through the chamber to label the nuclei of the cells. The sample was then washed immediately with 2 mL of 2xSSC followed by 2 mL of imaging buffer. Each region of the sample was then imaged once again with ~1 mW of 405-nm light.

Because cells were imaged using wide-field imaging with oblique-incidence illumination, without optical sectioning and z-scanning, the fraction of individual RNA species that was outside the axial range of the imaging geometry was quantified for 6 different RNA species using conventional smFISH. For this purpose, these cells were optically sectioned by collecting stacks of images at different focal depths through the entire depth of the cells. The images were aligned in consecutive focal planes and then computed for each cell the fraction of RNAs that were detected in the three-dimensional stack but not in the basal focal plane. It was found that only a small fraction, 15% +/- 1% (Mean +/- SEM across six different RNA species) of RNA molecules were outside the imaging range of a fixed focal plane without z-scanning. These measurements also confirmed that the excitation geometry illuminated the full depth of the cells. Any optical sectioning technique could be employed in MERFISH to allow the imaging of RNAs in thicker cells or tissues.

Construction of measured words. Fluorescent spots were identified and localized in each image using a multi-Gaussian-fitting algorithm assuming a Gaussian with a uniform width of 167 nm. This algorithm was used to allow partially overlapping spots to be distinguished and individually fit. RNA spots were distinguished from background signal, i.e. signal arising from probes bound non-specifically, by setting the intensity threshold

required to fit a spot with this software. Due to variation in the brightness of spots between rounds of hybridization, this threshold was adjusted appropriately for each hybridization round to minimize the combined average of the 1-->0 and 0-->1 error rates across all hybridization rounds (140-gene measurements) or to maximize the ratio of the number of measured words with four '1' bits to those with three or five '1' bits (1001-gene measurements). The location of the fiducial beads was identified in each frame using a faster single-Gaussian fitting algorithm.

Images of the same sample region in different rounds of hybridization were registered by rotating and translating the image to align the two fiducial beads within the same image that were most similar in location after a coarse initial alignment via image correlation. All images were aligned to a coordinate system established by the images collected in the first round of hybridization. The quality of this alignment was determined from the residual distance between five additional fiducial beads, and alignment error was typically ~20 nm.

Fluorescence spots in different hybridization rounds were connected into a single string, corresponding to a potential RNA molecule, if the distance between spots was smaller than 1 pixel (167 nm). For each string of spots, the on-off sequence of fluorescent signals in all hybridization rounds were used to assign a binary word to the potential RNA molecule, in which '1' was assigned to the hybridization rounds that contained a fluorescent signal above threshold and '0' was assigned to the other hybridization rounds. Measured words were then decoded into RNA species using the 16-bit MHD4 code or the 14-bit MHD2 code discussed above. In the case of the 16-bit MHD4 code, if the measured binary word matched the code word of a specific RNA perfectly or differed from the code word by one single bit, it was assigned to that RNA. In the case of the 14-bit MHD2 code, only if the measured binary word matched the code word of a specific RNA perfectly, was it assigned to that RNA. To determine the copy number per cell, the number of each RNA species was counted in individual cells within each 40 micrometer by 40 micrometer imaging area. It is noted that this number accounts for the majority but not all RNA molecules within a cell because a fraction of the cell could be outside the imaging area or focal depth. Tiling images of adjacent areas and adjacent focal planes could be employed to improve the counting accuracy.

In the 140-gene experiments, some regions of the cell nucleus occasionally contained too much fluorescence signal to properly identify individual RNA spots. In the 1001-gene experiments, the cell nucleus generally contained too much fluorescent signal to allow identification of individual RNA molecules. These bright regions were excluded from all

subsequent analysis. This work focuses on mRNAs, which are enriched in the cytoplasm. To estimate the fraction of mRNAs missed by excluding the nucleus region, conventional smFISH was used to quantify the fraction of molecules found inside the nucleus for six different mRNAs species. It was found that only 5% +/- 2% (Mean +/- SEM across six RNA species) of these RNA molecules are found in the nucleus. Employment of super-resolution imaging and/or optical sectioning could potentially allow individual molecules in these dense nucleus regions to be identified, which would be particularly useful for probing those non-coding RNAs that are enriched in the nucleus.

smFISH measurements of individual genes. Pools of 48 fluorescently-labeled (Quasar 670) oligonucleotide probes per RNA were purchased from Biosearch Technologies. 30-nt probe sequences were taken directly from a random subset of the targeting regions used for the multiplexed measurements. Cells were fixed and permeabilized as described above. 10 microliters of 250 nM oligonucleotide probes in encoding hybridization buffer (described above) was added to the cell-containing coverslip and spread uniformly by placing another coverslip on top of the sample. Samples were then incubated in a humid chamber inside a 37 ^{o}C-hybridization oven for 18 hours. Cells were then washed with encoding wash buffer (described above) at 37 ^{o}C for 10 minutes, and this wash was repeated for a total of three times. The sample was then washed three times with 2xSSC and imaged in imaging buffer using the same imaging geometry as described above for MERFISH.

Bulk RNA sequencing. Total RNA was extracted from IMR90 cells cultured as above using the Zymo Quick RNA MiniPrep kit (R1054) according to the manufacturer's instructions. polyA RNA was then selected (NEB; E7490), and a sequencing library was constructed using the NEBNext Ultra RNA library preparation kit (NEB; E7530), amplified with custom oligonucleotides, and 150-bp reads were obtained from on a MiSeq. These sequences were aligned to the human genome (Gencode v18) and isoform abundance was computed with cufflinks.

Calculation of the predicted scaling and error properties of different encoding schemes. Analytic expressions were derived for the dependence of the number of possible code words, the calling rate, and the misidentification rate on N. The calling rate is defined as the fraction of RNA molecules that are properly identified. The misidentification rate is defined as the fraction of RNA molecules that are misidentified as a wrong RNA species. For encoding schemes with an error-detection capability, the calling rate and misidentification rate does not add up to 1 because a fraction of the molecules not called properly can be detected as errors and discarded and, hence, not misidentified as a wrong

species. These calculations assume that the probability of misreading bits is constant for all hybridization rounds but differs for the 1-->0 and 0-->1 errors. Experimentally measured average 1-->0 and 0-->1 error rates (10% and 4% respectively) were used for the estimates shown in Figs. 5B-5D. For simplicity, the word corresponding to all '0's was not removed from calculations.

For the simple binary encoding scheme in which all possible *N*-bit binary words are assigned to unique RNA species, the number of possible code words is 2^N. The number of words that could be used to encode RNA is actually $2^N - 1$ because the code word '00…0' does not contain detectable fluorescence in any hybridization round, but for simplicity the word corresponding to all '0's was not removed from subsequent calculations. The error introduced by this approximation is negligible. For any given word with *m* '1's and *N–m* '0's the probability of measuring that word without error (the fraction of RNAs that is properly called) is:

$$(1-p_1)^m(1-p_0)^{N-m}, \tag{1}$$

where p_1 is 1-->0 error rate and p_0 is 0-->1 error rate per bit. Because different words in this simple binary encoding scheme can have different numbers of '1' bits, the calling rate for different words will differ if $p_1 \neq p_0$. The average calling rate, reported in Fig. 5C, was determined from the weighted average of the value of Eq. (1) for all words. This weighted average is:

$$\frac{1}{2^N}\sum_{m=0}^{N}\binom{N}{m}(1-p_1)^m(1-p_0)^{N-m}, \tag{2}$$

where $\binom{N}{m}$ is the binomial coefficient and corresponds to the number of words with *m* '1' bits in this encoding scheme. Since in this encoding scheme every error produces a binary word that encodes a different RNA, the average misidentification rate for this encoding scheme, reported in Fig. 5D, follows directly from (2):

$$1-\frac{1}{2^N}\sum_{m=0}^{N}\binom{N}{m}(1-p_1)^m(1-p_0)^{N-m}. \tag{3}$$

To calculate the scaling and error properties of the extended Hamming distance 4 (HD4) code, the generator matrix for the desired number of data bits using standard methods was first created. The generator matrix determines the specific words that are present in a given encoding scheme and was used to directly determine the number of encoded words as a function of the number of bits. In this encoding scheme, the calling rate corresponds to the

fraction of words measured without error as well as the fraction of words measured with a single-bit error. For code words with m '1' bits, this fraction is determined by the following expression:

$$(1-p_1)^m(1-p_0)^{N-m} + m\,p_1{}^1(1-p_1)^{m-1}(1-p_0)^{N-m} + (N-m)\,p_0{}^1(1-p_1)^m(1-p_0)^{N-m-1} \quad (4)$$

where the first term is the probability of not making any errors, the second term corresponds to the total probability of making one 1-->0 error at any of the m '1' bits without making any other 0-->1 errors, and the final term corresponds to the total probability of making one 0-->1 error at any of the N-m '0' bits without making any 1-->0 errors. Because the number of '1' bits can differ between words in this encoding scheme, the average calling rate reported in Fig. 5C was computed from a weighted average over Eq. (4) for different values of m. The weight for each term was determined from the number of words that contain m '1' bits as determined from the generator matrix described above.

Because RNA-encoding words are separated by a minimum Hamming distance of 4, at least 4 errors are required to switch one word into another. If error correction is applied, then 3 or 5 errors could also convert one RNA into another. Thus, the misidentification rate from all possible combinations of 3-bit, 4-bit and 5-bit errors was estimated for code words with m '1' bits. Technically, >5-bit errors could also convert one RNA into another, but the probability of making such errors is negligible because of the small per-bit error rate. This expression was approximated with:

$$\begin{aligned}
&\sum_{i=0}^{4}\binom{m}{i}\binom{N-m}{4-i}p_1^{\,i}p_0^{\,4-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(4-i)} + \\
&\sum_{i=0}^{3}\binom{m}{i}\binom{N-m}{3-i}p_1^{\,i}p_0^{\,3-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(3-i)} + \\
&\sum_{i=0}^{5}\binom{m}{i}\binom{N-m}{5-i}p_1^{\,i}p_0^{\,5-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(5-i)} +
\end{aligned} \quad (5)$$

The first sum corresponds to all of the ways in which exactly four mistakes can be made. Similarly, the second and third sums correspond to all of the ways in which exactly three or five mistakes can be made. Eq. (5) provides an upper bound for the misidentification rate because not all three, four, or five bit errors produce a word that matches or would be corrected to another legitimate word. Again because the number of '1' bits can differ between words, the average misidentification rate reported in Fig. 5D is calculated as a weighted average of Eq. (5) over the number of words that have m '1' bits.

To generate the MHD4 code where the number of '1' bits for each code word is set to 4, the HD4 codes were first generated as described above, and then all code words that did

not contain four '1's were removed. The calling rate of this code, reported in Fig. 5C, was directly calculated from Eq. (4) but with m = 4 because all code words in this code have four '1' bits. The misidentification rate of this code, reported in Fig. 5D, was calculated by modifying Eq. (5) with the following considerations: (i) the number of '1' bits, m, was set to 4 and (ii) errors that produce words that do not contain three, four, or five '1' bits were excluded. Thus, the expression in Eq. (5) was simplified to

$$\begin{aligned}&\binom{4}{2}\binom{N-4}{2}p_1^2p_0^2(1-p_1)^2(1-p_0)^{N-6}+\\&\binom{4}{1}\binom{N-4}{2}p_1^1p_0^2(1-p_1)^3(1-p_0)^{N-6}+\binom{4}{2}\binom{N-4}{1}p_1^2p_0^1(1-p_1)^2(1-p_0)^{N-5}\\&\binom{4}{2}\binom{N-4}{3}p_1^2p_0^3(1-p_1)^2(1-p_0)^{N-7}+\binom{4}{3}\binom{N-4}{2}p_1^3p_0^2(1-p_1)^1(1-p_0)^{N-6}\end{aligned} \quad (6)$$

Again, this expression is an upper bound on the actual misidentification rate because not all words with four '1's are valid code words.

Estimates of the 1-->0 and 0-->1 error rates for each hybridization round. To compute the probability of misreading a bit at a given hybridization round, the error correcting properties of the MHD4 code were used. Briefly, the probabilities of 1-->0 or 0-->1 errors were derived in the following way. Let the probability of making an error at the *i*th bit, i.e. *i*th hybridization round, be p_i and the actual number of RNA molecules of the given species be A, then the number of exact matches for this RNA will be $W_E = A\prod_{i=1}^{16}(1-p_i)$ and the number of one-bit error corrected matches for this RNA corresponding to errors at the *i*th bit will be $W_i = A\frac{p_i}{(1-p_i)}\prod_{j=1}^{16}(1-p_j)$. The p_i can be directly derived from the ratio: $W_i/W_E = \frac{p_i}{(1-p_i)}$. This ratio assumes that the one-bit error-corrected counts were only generated from single-bit errors from the correct word and that multi-error contamination from other RNA words is negligible. Given that the error rate per hybridization round is small and that it takes at least three errors to convert one RNA-encoding word into a word that would be misidentified as another RNA, the above approximation should be a good one.

To compute the average 1-->0 or 0-->1 error probabilities for each of the 16 hybridization rounds, the above approach was used to calculate the per-bit error rates for each bit of every gene, and these errors were sorted based on whether they correspond to a 1-->0 or a 0-->1 error, and the average of these errors for each bit weighted by the number of counts observed for the corresponding gene was taken.

Estimates of the calling rate for individual RNA species from actual imaging data. With the estimates of the 1-->0 or 0-->1 error probabilities for each round of hybridization as determined above, it is possible to estimate the calling rate for each RNA based on the specific word used to encoded it. Specifically, the fraction of an RNA species that is called correctly is determined by

$$\prod_{i=1}^{N}(1-p_i)+\sum_{j=1}^{N}\frac{p_j}{(1-p_j)}\prod_{i=1}^{N}(1-p_i), \quad (7)$$

where the first term represent the probability of observing an exact match of the code word and the second term represent the probability of observing an error-corrected match (i.e. with one-bit error). The values of the per-bit error rate p_i for each RNA species are determined by the specific code word for that RNA and the measured 1-->0 or 0-->1 error rates for each round of hybridization. If the code word of the RNA contains a '1' in the *i*th bit, then p_i is determined from the 1-->0 error rate for the *i*th hybridization round; if the word contains a '0' in the *i*th bit, p_i is determined from the 0-->1 error rate for the *i*th hybridization round.

Hierarchical clustering analysis of the co-variation in RNA abundance. Hierarchical clustering of the co-variation in gene expression for both the 140-gene and 1001-gene experiments was conducted as follows. First, the distance between every pair of genes was determined as 1 minus the Pearson correlation coefficient of the cell-to-cell variation of the measured copy numbers of these two RNA species, both normalized by the total RNA counted in the cell. Thus, highly correlated genes are 'closer' to one another and highly anti-correlated genes are 'further' apart. An agglomerative hierarchical cluster tree was then constructed from these distances using the Unweighted Pair Group Method with Arithmetic mean (UPGMA). Specifically, starting with individual genes, hierarchical clusters were constructed by identifying the two clusters (or individual genes) that are closest to one another according to the arithmetic mean of the distances between all inter-cluster gene pairs. The pairs of clusters (or individual genes) with the smallest distance are then grouped together and the process is repeated. The matrix of pairwise correlations was then sorted based on the order of the genes within these trees.

Groups of genes with substantial co-variations were identified by selecting a threshold on the hierarchical cluster tree (indicated by the dashed lines in Figs. 7D and 10A) that produced approximately 10 groups of genes each of which contains at least 4 members for the 140-gene experiments or approximately 100 groups each of which contains at least 3 members for the 1001-gene experiments. It is noted that one can change the threshold in

order to identify either more tightly coupled smaller groups or larger groups with relatively loose coupling.

A probability value for the confidence that a gene belongs to a specific group was determined by computing the difference between the average correlation coefficient between that gene and all other members of that group and the average correlation coefficient between that gene and all other measured genes outside that group. The significance (p-value) of this difference was determined with the student's t-test.

Because hierarchical clustering is inherently a one-dimensional analysis, i.e. any given genes can only be a member of a single group, this analysis does not allow all correlated gene groups to be identified. Higher dimension analysis, such as principal component analysis or k-means clustering, could be used to identify more co-varying gene clusters.

Analysis of RNA spatial distributions. To identify genes that have similar spatial distributions, each of the measured cells was subdivided into 2x2 regions and calculated the fraction of each RNA species present in each of these bins. To control for the fact that some regions of the cell naturally contain more RNA than others, the enrichment was calculated for each gene , i.e., the ratio of the observed fraction in a given region for a given RNA species to the average fraction observed for all genes in that same region. For each pair of RNA species, the Pearson correlation coefficient of the region-to-region variation in enrichment of these two RNA species for each cell was determined and the correlation coefficients were averaged over ~400 cells imaged in 7 independent data sets. RNA species were then clustered based on these average correlation coefficients using the same hierarchical clustering algorithm described above. Because of the large number of cells used for the analysis, it was found that the coarse spatial binning (2x2 regions per cell) was sufficient to capture the spatial correlation between genes and finer binning did not produce more significantly correlated groups.

To measure the distances of genes from the nuclei and from the cell edge, brightness thresholds on the cell images were first used to segment the nuclei and the cell edges identified. The distance from every RNA molecule to the nearest part of the nucleus and nearest part of the cell edge was then determined. For each data set, the average distance for each RNA species averaged over all the cells measured was then computed. These distances were averaged for the group I genes, group II genes or all genes. Only those RNA species with at least 10 counts per cell were used in this analysis to minimize statistical error on the distance values.

Gene ontology (GO) analysis. Groups of genes were selected from the hierarchical trees as discussed above. A collection of GO terms was determined for all measured RNA species as well as the RNA species associated with each group from the most recent human GO annotations using both the annotated GO terms and terms immediately upstream or downstream of the found annotations. The enrichment of these annotations was calculated from the ratio of the fraction of genes within each group that have this term to the fraction of all measured genes that have this term and the p-value for this enrichment was calculated via the hypergeometric function. Only statistically significantly enriched GO terms with a p-value less than 0.05 were considered.

While several embodiments of the present invention have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the functions and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the present invention. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the teachings of the present invention is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, the invention may be practiced otherwise than as specifically described and claimed. The present invention is directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles, materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the scope of the present invention.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

The indefinite articles “a” and “an,” as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean “at least one.”

The phrase "and/or," as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with "and/or" should be construed in the same fashion, i.e., "one or more" of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the "and/or" clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to "A and/or B", when used in conjunction with open-ended language such as "comprising" can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, "or" should be understood to have the same meaning as "and/or" as defined above. For example, when separating items in a list, "or" or "and/or" shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as "only one of" or "exactly one of," or, when used in the claims, "consisting of," will refer to the inclusion of exactly one element of a number or list of elements. In general, the term "or" as used herein shall only be interpreted as indicating exclusive alternatives (i.e. "one or the other but not both") when preceded by terms of exclusivity, such as "either," "one of," "only one of," or "exactly one of." "Consisting essentially of," when used in the claims, shall have its ordinary meaning as used in the field of patent law.

As used herein in the specification and in the claims, the phrase "at least one," in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase "at least one" refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, "at least one of A and B" (or, equivalently, "at least one of A or B," or, equivalently "at least one of A and/or B") can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and

optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

When the word "about" is used herein in reference to a number, it should be understood that still another embodiment of the invention includes that number not modified by the presence of the word "about."

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

In the claims, as well as in the specification above, all transitional phrases such as "comprising," "including," "carrying," "having," "containing," "involving," "holding," "composed of," and the like are to be understood to be open-ended, i.e., to mean including but not limited to. Only the transitional phrases "consisting of" and "consisting essentially of" shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

CLAIMS

1. A method, comprising:

exposing a sample to a plurality of nucleic acid probes;

for each of the nucleic acid probes, determining binding of the nucleic acid probes within the sample;

creating codewords based on the binding of the nucleic acid probes; and

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

2. The method of claim 1, comprising exposing the sample to at least 5 different nucleic acid probes.

3. The method of any one of claims 1 or 2, comprising exposing the sample to at least 10 different nucleic acid probes.

4. The method of any one of claims 1-3, comprising exposing the sample to at least 100 different nucleic acid probes.

5. The method of any one of claims 1-4, comprising exposing the sample to a plurality of nucleic acid probes simultaneously.

6. The method of any one of claims 1-5, comprising exposing the sample to a plurality of nucleic acid probes sequentially.

7. The method of claims 1-6, wherein the plurality of nucleic acid probes comprises a combinatorial combination of nucleic acid probes with different sequences.

8. The method of claim 7, wherein the combinatorial combination of nucleic acid probes target a combinatorial combination of RNA species in the sample.

9. The method of claim 7 or 8, wherein the combinatorial combination of nucleic acid probes target a combinatorial combination of DNA sequences in the sample.

10. The method of any one of claims 1-9, wherein at least some of the plurality of nucleic acid probes comprises a first portion comprising a target sequence and a second portion comprising one or more read sequences.

11. The method of claim 10, wherein the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from the one or more read sequences.

12. The method of any one of claims 10 or 11, wherein the target sequence is substantially complementary to a nucleic acid sequence encoding a protein.

13. The method of any one of claims 10-12, wherein the target sequence is substantially complementary to messenger RNA (mRNA).

14. The method of any one of claims 10-13, wherein the plurality of nucleic acid probes comprises at least 8 possible read sequences.

15. The method of any one of claims 10-14, wherein the plurality of nucleic acid probes comprises no more than 32 possible read sequences.

16. The method of any one of claims 10-15, wherein the plurality of nucleic acid probes comprises no more than 16 possible read sequences.

17. The method of any one of claims 10-16, wherein the plurality of nucleic acid probes comprises no more than 8 possible read sequences.

18. The method of any one of claims 10-17, wherein the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

19. The method of any one of claims 10-18, wherein the target sequence of the plurality of nucleic acid probes has an average length of between 10 and 200 nucleotides.

20. The method of any one of claims 10-19, wherein the plurality of read sequences have an average length of between 5 nucleotides and 50 nucleotides.

21. The method of any one of claims 10-20, wherein at least some of the plurality of nucleic acid probes comprise no more than 10 read sequences.

22. The method of any one of claims 10-21, wherein at least some of the plurality of nucleic acid probes comprise no more than 6 read sequences.

23. The method of any one of claims 10-22, wherein at least some of the plurality of nucleic acid probes comprise no more than 4 read sequences.

24. The method of any one of claims 10-23, wherein at least some of the plurality of nucleic acid probes comprise no more than 3 read sequences.

25. The method of any one of claims 10-24, wherein at least some of the plurality of nucleic acid probes comprise no more than 2 read sequences.

26. The method of claims 1-25, further comprising exposing the sample to a first secondary probe comprising a first signaling entity, the first secondary probe able to bind to some of the read sequences of the nucleic acid probes, and determining binding of the nucleic acid probes by determining the first signaling entity within the sample.

27. The method of claim 26, wherein the first signaling entity is fluorescent.

28. The method of any one of claims 26 or 27, wherein the first signaling entity is a protein.

29. The method of any one of claims 26-28, wherein the first signaling entity is a dye.

30. The method of any one of claims 26-29, wherein the first signaling entity is a nanoparticle.

31. The method of any one of claims 26-30, further comprising exposing the sample to a second secondary probe comprising a second signaling entity, the second secondary probe able to bind to some of read sequences of the nucleic acid probes, and determining binding of the nucleic acid probes by determining the second signaling entity within the sample.

32. The method of claim 31, wherein the first signaling entity and the second signaling entity are identical.

33. The method of claim 31, wherein the first signaling entity and the second signaling entity are not identical.

34. The method of any one of claims 31-33, further comprising inactivating the first signaling entity prior to exposing the sample to the second secondary probe.

35. The method of claim 34, comprising inactivating the first signaling entity by photobleaching at least some of the first signaling entity.

36. The method of any one of claims 34 or 35, comprising inactivating the first signaling entity by chemically bleaching at least some of the first signaling entity.

37. The method of any one of claims 34-36, comprising inactivating the first signaling entity by exposing the first signaling entity to a reactant able to alter the structure of the signaling entity.

38. The method of any one of claims 34-37, comprising inactivating the first signaling entity by removing at least some of the first signaling entity.

39. The method of any one of claims 34-38, comprising inactivating the first signaling entity by dissociating the first signaling entity from the first secondary probe.

40. The method of any one of claims 34-39, comprising inactivating the first signaling entity by dissociating the first secondary probe that contains the first signaling entity from the sample.

41. The method of any one of the claims 34-40, comprising inactivating the first signaling entity by chemically cleaving it from the first secondary probe.

42. The method of any one of the claims 34-41, comprising inactivating the first signaling entity by enzymatically cleaving it from the first secondary probe.

43. The method of any one of claims 34-42, comprising inactivating the first signaling entity by exposing the signaling entity or the first secondary probe to an enzyme.

44. The method of any one of claims 1-43, comprising determining the centroid of the first signaling entity using an algorithm for determining non-overlapping single emitters.

45. The method of any one of claims 1-44, comprising determining the centroid of the first signaling entity using an algorithm for determining partially overlapping single emitters.

46. The method of any one of claims 44 or 45, comprising determining the centroid using a maximum likelihood algorithm.

47. The method of any one of claims 44-46, comprising determining the centroid using a least squares algorithm.

48. The method of any one of claims 44-47, comprising determining the centroid using a Bayesian algorithm.

49. The method of any one of claims 44-48, comprising determining the centroid using a compressed sensing algorithm.

50. The method of any one of claims 1-49, further comprising determining a confidence level for the identified nucleic acid target.

51. The method of claim 51, comprising determining the confidence level using a ratio of the number of exact matches to the number of matches having one or more one-bit errors to the codeword.

52. The method of claim 51, comprising determining the confidence level using a ratio of the number of exact matches to the number of matches having exactly one one-bit error to the codeword.

53. The method of claims 1-52, wherein at least some of the plurality of nucleic acid probes comprise DNA.

54. The method of any one of claims 1-53, wherein at least some of the plurality of nucleic acid probes comprise RNA.

55. The method of any one of claims 1-54, wherein at least some of the plurality of nucleic acid probes comprise PNA.

56. The method of any one of claims 1-55, wherein at least some of the plurality of nucleic acid probes comprise LNA.

57. The method of any one of claims 1-56, wherein the plurality of nucleic acid probes have an average length of between 10 and 300 nucleotides.

58. The method of any one of claims 1-57, wherein at least some of the plurality of nucleic acids probes are configured to bind to a nucleic acid within the sample.

59. The method of any one of claims 1-58, wherein at least some of the binding of the nucleic acid probes to a target within the sample is specific binding.

60. The method of any one of claims 1-59, wherein at least some of the binding of the nucleic acid probes to a target within the sample is via Watson-Crick base pairing.

61. The method of any one of claims 1-60, wherein at least some of the plurality of nucleic acids probes are configured to bind to RNA.

62. The method of any one of claims 1-61, wherein at least some of the plurality of nucleic acid probes are configured to bind to non-coding RNA.

63. The method of any one of claims 1-62, wherein at least some of the plurality of nucleic acid probes are configured to bind to mRNA.

64. The method of any one of claims 1-63, wherein at least some of the plurality of nucleic acid probes are configured to bind to transfer RNA (tRNA).

65. The method of any one of claims 1-64, wherein at least some of the plurality of nucleic acid probes are configured to bind to ribosomal RNA (rRNA).

66. The method of any one of claims 1-65, wherein at least some of the plurality of nucleic acid probes are configured to bind to lncRNA.

67. The method of any one of claims 1-66, wherein at least some of the plurality of nucleic acid probes are configured to bind to snoRNA.

68. The method of any one of claims 1-67, wherein at least some of the plurality of nucleic acid probes are configured to bind to non-coding RNA.

69. The method of any one of claims 1-68, wherein the plurality of nucleic acids probes is configured to bind to DNA.

70. The method of any one of claims 1-69, wherein at least some of the plurality of nucleic acid probes are configured to bind to genomic DNA.

71. The method of any one of claims 1-70, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 300 nm.

72. The method of any one of claims 1-71, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 100 nm.

73. The method of any one of claims 1-72, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 80 nm.

74. The method of any one of claims 1-73, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 50 nm.

75. The method of any one of claims 1-74, wherein the sample comprises a cell.

76. The method of claim 75, wherein the cell is a human cell.

77. The method of any one of claims 76 or 77, wherein the cell is fixed.

78. The method of any one of claims 1-77, comprising determining binding of the nucleic acid probes by imaging at least a portion of the sample.

79. The method of any one of claims 1-78, comprising determining binding of the nucleic acid probes using an optical imaging technique.

80. The method of any one of claims 1-79, comprising determining binding of the nucleic acid probes using a fluorescence imaging technique.

81. The method of any one of claims 1-80, comprising determining binding of the nucleic acid probes using a multi-color fluorescence imaging technique.

82. The method of any one of claims 1-81, comprising determining binding of the nucleic acid probes using a super-resolution fluorescence imaging technique.

83. The method of claim 82, comprising determining binding of the nucleic acid probes using stochastic optical reconstruction microscopy (STORM).

84. The method of any one of claims 82 or 83, comprising determining binding of the nucleic acid probes using photoactivated localization microscopy (PALM) or fluorescence photoactivation localization microscopy (FPALM).

85. The method of any one of claims 82-84, comprising determining binding of the nucleic acid probes using stimulated emission depletion microscopy (STED).

86. The method of any one of claims 82-85, comprising determining binding of the nucleic acid probes using structured illumination microscopy (SIM).

87. The method of any one of claims 82-86, comprising determining binding of the nucleic acid probes using reversible saturated optically linear fluorescence transition (RESOLFT) microscopy.

88. A method, comprising:

exposing a sample to a plurality of nucleic acid probes, wherein the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences, and wherein at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences; and

for each of the nucleic acid probes, determining binding of the target sequences of the nucleic acid probes within the sample.

89. The method of claim 88, further comprising:

creating codewords based on determination of the read sequences within the sample.

90. The method of claim 89, further comprising:

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

91. A method, comprising:

exposing a sample to a plurality of primary nucleic acid probes;

exposing the plurality of primary nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary nucleic acid probes within the sample;

creating codewords based on fluorescence of the secondary nucleic acid probes; and

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

92. A method, comprising:

exposing a plurality of primary nucleic acid probes to a sample; and

exposing the plurality of nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary probes within the sample, wherein at least some of the plurality of secondary nucleic acid probes comprises distinguishable secondary nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences.

93. The method of claim 92, further comprising:

creating codewords based on fluorescence of the secondary nucleic acid probes.

94. The method of claim 93, further comprising:

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

95. A composition, comprising:

a plurality of nucleic acid probes, each comprising a first portion comprising a target sequence and a plurality of read sequences, wherein the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

96. The composition of claim 95, wherein the plurality of nucleic acid probes defines a code space with a Hamming distance of at least 2.

97. The composition of claim 96, wherein the plurality of nucleic acid probes defines a code space with a Hamming distance of at least 3.

98. The composition of any one of claims 96 or 97, wherein the code is a Hamming(7, 4) code, a Hamming(15, 11) code, a Hamming(31, 26) code, a Hamming(63, 57) code, or a Hamming(127, 120) code.

99. The composition of any one of claims 95-98, wherein the code is a SECDED code.

100. The composition of claim 99, wherein the code is a SECDED(8,4) code.

101. The composition of claim 99, wherein the code is a SECDED(16,4) code.

102. The composition of claim 99, wherein the code is a SECDED(16, 11) code, a SECDED (22, 16) code, a SECDED (39, 32) code, or a SECDED (72, 64) code.

103. The composition of any one of claims 99-102, wherein only codewords with a constant number of ones are used.

104. The composition of claim 99, wherein the code is MHD4 code.

105. The composition of claim 99, wherein the code is MHD2 code.

106. The composition of any one of claims 95-105, wherein the plurality of nucleic acid probes comprises no more than 100 possible read sequences.

107. The composition of any one of claims 95-106, wherein the plurality of nucleic acid probes comprises no more than 64 possible read sequences.

108. The composition of any one of claims 95-107, wherein the plurality of nucleic acid probes comprises no more than 32 possible read sequences.

109. The composition of any one of claims 95-108, wherein the plurality of nucleic acid probes comprises no more than 16 possible read sequences.

110. The composition of any one of claims 95-109, wherein the plurality of nucleic acid probes comprises no more than 8 possible read sequences.

111. The composition of any one of claims 95-110, wherein the target sequence of the plurality of nucleic acid probes has an average length of between 10 and 200 nucleotides.

112. The composition of any one of claims 95-111, wherein the plurality of read sequences have an average length of between 5 nucleotides and 50 nucleotides.

113. The composition of any one of claims 95-112, wherein at least some of the plurality of nucleic acid probes comprise no more than 10 read sequences.

114. The composition of any one of claims 95-113, wherein at least some of the plurality of nucleic acid probes comprise no more than 6 read sequences.

115. The composition of any one of claims 95-114, wherein at least some of the plurality of nucleic acid probes comprise no more than 4 read sequences.

116. The composition of any one of claims 95-115, wherein at least some of the plurality of nucleic acid probes comprise no more than 2 read sequences.

117. The composition of any one of claims 95-116, wherein at least some of the plurality of nucleic acid probes comprise DNA.

118. The composition of any one of claims 95-117, wherein at least some of the plurality of nucleic acid probes comprise RNA.

119. The composition of any one of claims 95-118, wherein at least some of the plurality of nucleic acid probes comprise PNA.

120. The composition of any one of claims 95-119, wherein at least some of the plurality of nucleic acid probes comprise LNA.

121. The composition of any one of claims 95-120, wherein the plurality of nucleic acid probes have an average length of between 10 and 300 nucleotides.

122. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords; and

forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

123. The method of claim 122, wherein the code space has a Hamming distance of at least 2.

124. The method of any one of claims 122 or 123, wherein the plurality of targets comprise RNA.

125. The method of any one of claims 122-124, wherein the plurality of targets comprise DNA.

126. A method, comprising:

exposing nucleic acid probes formed using the method of any one of claims 122-125 to a sample;

sequentially determining binding of each of a plurality of secondary nucleic probes to the nucleic acid probes within the sample;

determining sample codewords comprising a pattern of bindings of the plurality of secondary nucleic probes to the nucleic acid probes; and

matching the sample codewords to the codewords to determining targets

within the sample.

127. The method of claim 126, wherein at least some of the plurality of secondary nucleic probes is able to bind to a read sequence of a nucleic acid probe.

128. The method of claim 127, wherein at least some of the plurality of secondary nucleic probes is able to specifically bind to only one read sequence.

129. The method of claim 125-128, wherein if no match is found between a valid codeword and the sample codeword, applying error correction to the sample codeword to identified a matched codeword.

130. The method of claim 125-129, wherein at least some of the secondary nucleic acid probes each comprise a signaling entity.

131. The method of claim 130, comprising imaging the signaling entity to determine binding of the plurality of secondary nucleic probes within the sample.

132. The method of claim 131, comprising imaging the signaling entity using a super-resolution imaging technique.

133. The method of any one of claims 131 or 132, further comprising applying drift correction to the images.

134. The method of any one of claims 131-133, further comprising applying physical registration to the images.

135. The method of claim 125-134, comprising determining binding of the plurality of secondary nucleic probes within the sample at a resolution better than 100 nm.

136. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-

correcting code space;

forming a plurality of nucleic acid probes each comprising a target sequence; and

forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes corresponds to at least one common value of a position within the codewords.

137. The method of claim 136, where in the codewords comprise a number of positions that is less than the number of targets.

138. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions that is less than the number of targets, and wherein each codeword is associated with a single target;

associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords; and

forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

139. The method of claim 138, wherein the codewords form an error-checking space.

140. The method of any one of claims 138 or 139, wherein the codewords form an error-correcting code space.

141. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

forming a plurality of nucleic acid probes each comprising a target sequence; and

forming groups comprising the plurality of nucleic acid probes such that each

group of nucleic acid probes correspond to at least one common value of a position within the codewords.

142. The method of any one of claims 138-141, wherein the groups are applied to the sample sequentially using fluidic devices.

143. The method of any one of claims 138-142, wherein each group of nucleic acid probes contains a group identification sequence distinguishable from the group identification sequence of the other groups.

144. A method, comprising:

associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code;

associating unique read sequences with each possible value of each position in the codewords, wherein the read sequences are taken from a set of orthogonal sequences, which have limited homology with one another and with the nucleic acid species in a sample;

forming a plurality of primary nucleic acid probes each comprising a target sequence that uniquely binds to a nucleic acid target and one or more read sequences;

forming a plurality of secondary nucleic acid probes comprising a signaling entity and a sequence that is complementary to one of the read sequences;

exposing a sample to the primary nucleic acid probes such that the nucleic acid probes hybridize to the nucleic acid targets in the sample;

exposing the primary nucleic acid probes in the sample to a secondary nucleic acid probe such that the secondary nucleic acid probe hybridizes to the read sequence on at least some of the primary nucleic acid probes;

imaging the sample; and

repeating the exposing and imaging steps one or more times, using a different secondary nucleic acid probe for at least some of the repetitions.

145. The method of claim 144, wherein the wherein the read sequences are taken from a set of orthogonal sequences, which have a homology of less than 15 basepairs with

one another and with the nucleic acid species in a sample.

146. The method of any one of claims 144 or 145, comprising combining the images of the sample into a combined image.

147. The method of claim 146, further comprising identifying foreground features corresponding to putative nucleic acid targets in the sample.

148. The method of any one of claims 144-147, comprising determining the centroid of the first signaling entity using an algorithm for determining non-overlapping single emitters.

149. The method of any one of claims 144-148, comprising determining the centroid of the first signaling entity using an algorithm for determining partially overlapping single emitters.

150. The method of any one of claims 144-149, comprising determining the centroid using a maximum likelihood algorithm.

151. The method of any one of claims 144-150, comprising determining the centroid using a least squares algorithm.

152. The method of any one of claims 144-151, comprising determining the centroid using a Bayesian algorithm.

153 The method of any one of claims 144-152, comprising determining the centroid using a compressed sensing algorithm.

154. The method of any one of claims 144-153, comprising determining the codeword corresponding to each foreground feature by determining binding patterns of the nucleic acid probe pools to the sample.

155. The method of claim 154, comprising matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a

valid codeword that correspond to a nucleic acid target

156. The method of any one of claims 144-155, comprising imaging the sample using fluorescence microscopy.

157. The method of any one of claims 144-156, comprising imaging the sample at a resolution better than 500 nm.

158. The method of any one of claims 144-157, comprising imaging the sample at a resolution better than 300 nm.

159. The method of any one of claims 144-158, comprising imaging the sample at a resolution better than 100 nm.

160. The method of any one of claims 144-159, comprising imaging the sample at a resolution better than 80 nm.

161. The method of any one of claims 144-160, comprising imaging the sample at a resolution better than 50 nm.

162. The method of any one of claims 144-161, comprising imaging the sample at a resolution better than 20 nm.

163. The method of any one of claims 144-162, comprising imaging the sample using a super-resolution fluorescence imaging technique.

164. The method of any one of claims 144-163, comprising imaging the sample using a technique selected from the group consisting of STORM, PALM, FPALM, STED, SIM, RESOLFT, SOFI or SPDM.

165. The method of any one of claims 144-164, where in the codewords comprise a number of positions that is less than the number of targets.

166. A method, comprising:

associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

forming a plurality of nucleic acid probes comprising a signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets;

grouping the nucleic acid probes into a plurality of probe pools, wherein each of the probe pools corresponds to a specific value of a unique position within the codewords;

exposing a sample to one of the probe pools;

imaging the sample; and

repeating the exposing and imaging steps one or more times, using a different probe pool for at least some of the repetitions.

167. The method of claim 166, comprising combining the images of the sample into a combined image.

168. The method of claim 167, further comprising identifying foreground features corresponding to putative nucleic acid targets in the sample.

169. The method of claim 168, comprising determining the codeword corresponding to each foreground feature by determining binding patterns of the nucleic acid probe pools to the sample.

170. The method of claim 169, comprising matching the determined codeword to a valid codeword wherein, if no match is found, applying error correction to the determined codeword to form a valid codeword that correspond to a nucleic acid target.

171. The method of any one of claims 166-170, comprising imaging the sample using fluorescence microscopy.

172. The method of any one of claims 166-171, comprising imaging the sample at a resolution better than 300 nm.

173. The method of any one of claims 166-172, comprising imaging the sample at a resolution better than 100 nm.

174. The method of any one of claims 166-173, comprising imaging the sample using a super-resolution fluorescence imaging technique.

175. The method of any one of claims 166-174, comprising imaging the sample using a technique selected from the group consisting of STORM, PALM, FPALM, STED, SIM, RESOLFT, SOFI or SPDM.

176. The method of any one of claims 166-175, wherein the codewords comprise a number of positions that is less than the number of targets.

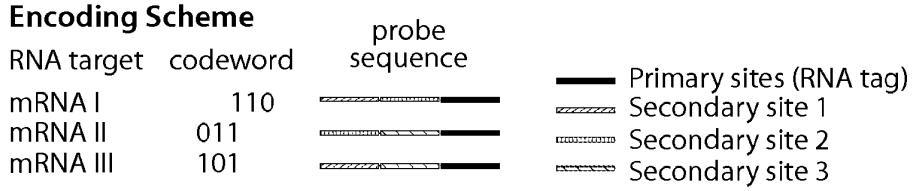


Fig. 1A

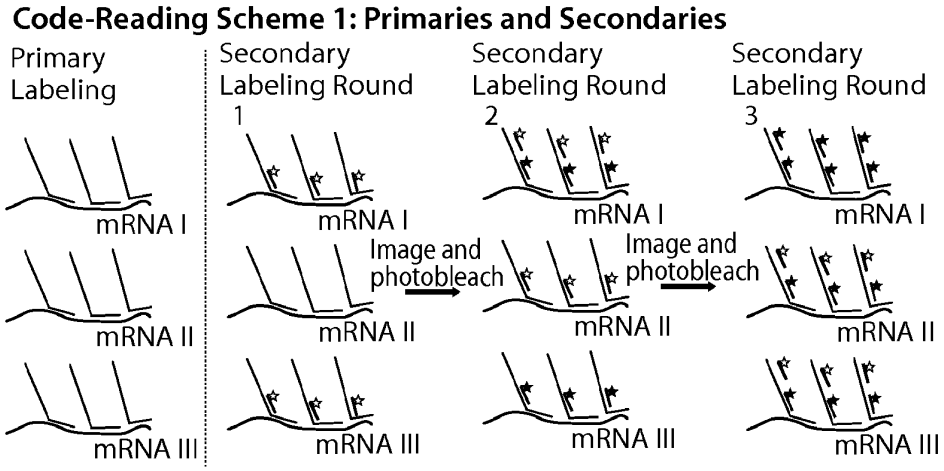


Fig. 1B

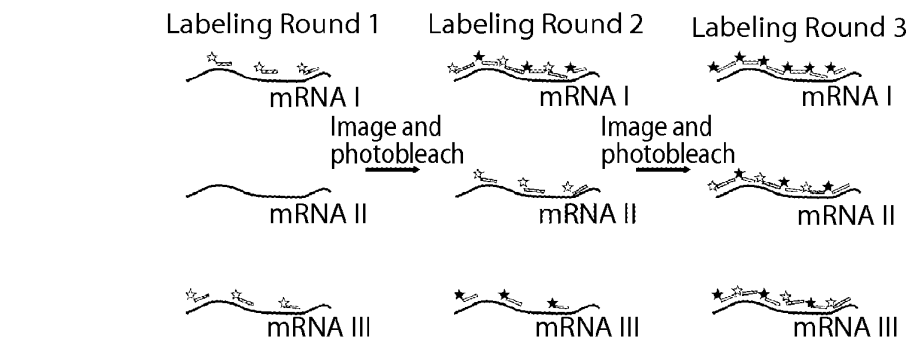


Fig. 1C

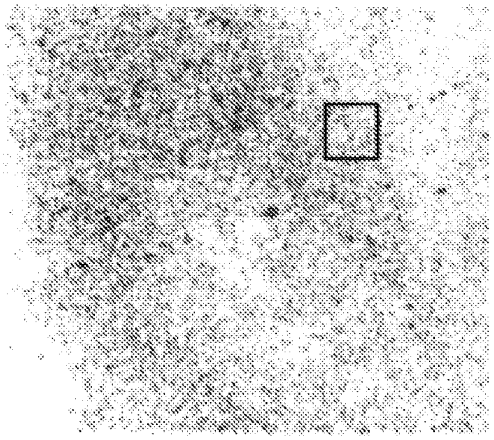


Fig. 2A

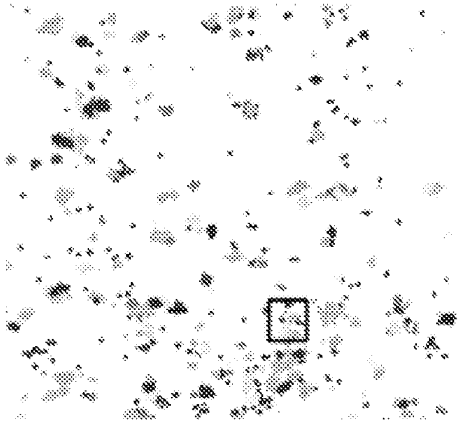


Fig. 2B



[0 1 0 1 1 1 0 0]

Fig. 2C

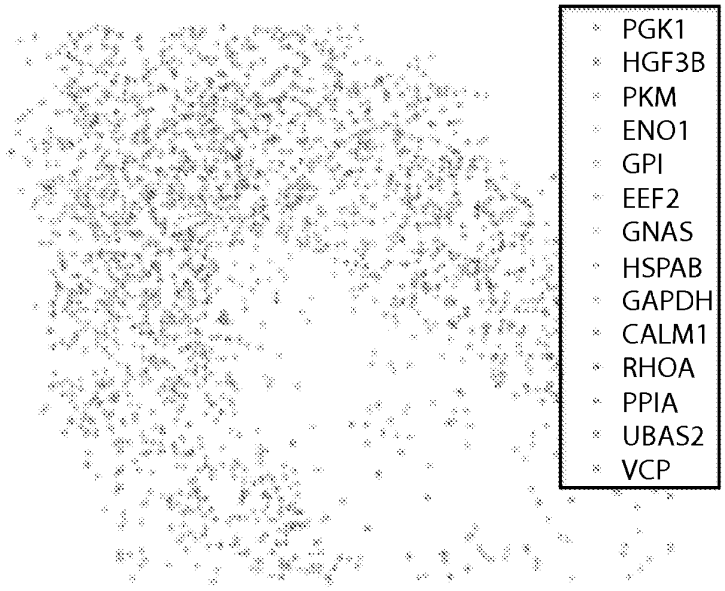

PGK1
HGF3B
PKM
ENO1
GPI
EEF2
GNAS
HSPAB
GAPDH
CALM1
RHOA
PPIA
UBAS2
VCP

Fig. 2D

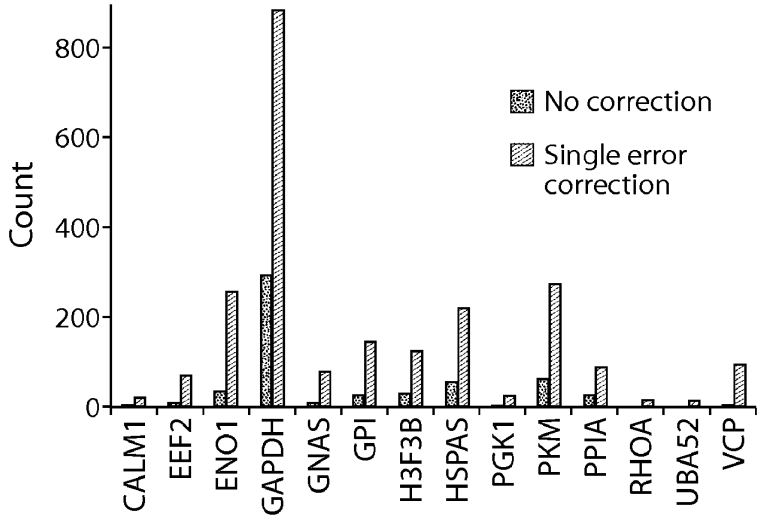

Count
800
600
400
200
0
No correction
Single error correction
CALM1
EEF2
ENO1
GAPDH
GNAS
GPI
H3F3B
HSPAS
PGK1
PKM
PPIA
RHOA
UBA52
VCP

Fig. 2E

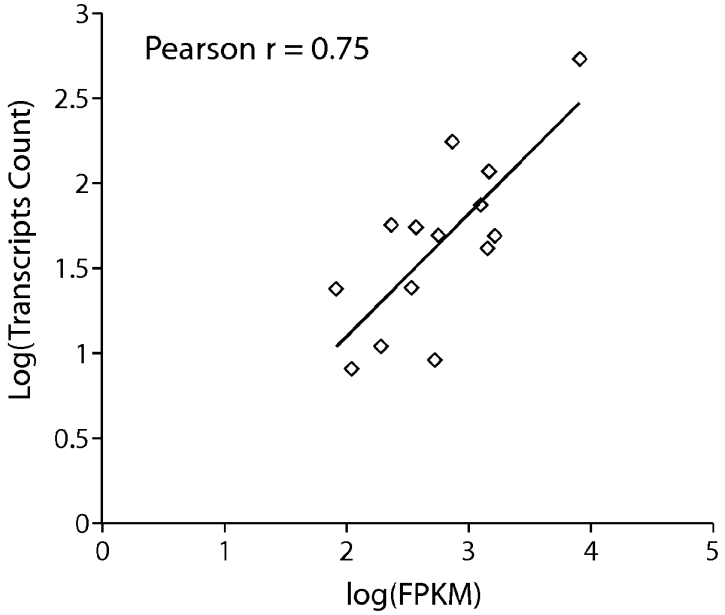

Pearson r = 0.75
Log(Transcripts Count)
log(FPKM)
0
0.5
1
1.5
2
2.5
3
0
1
2
3
4
5

Fig. 2F

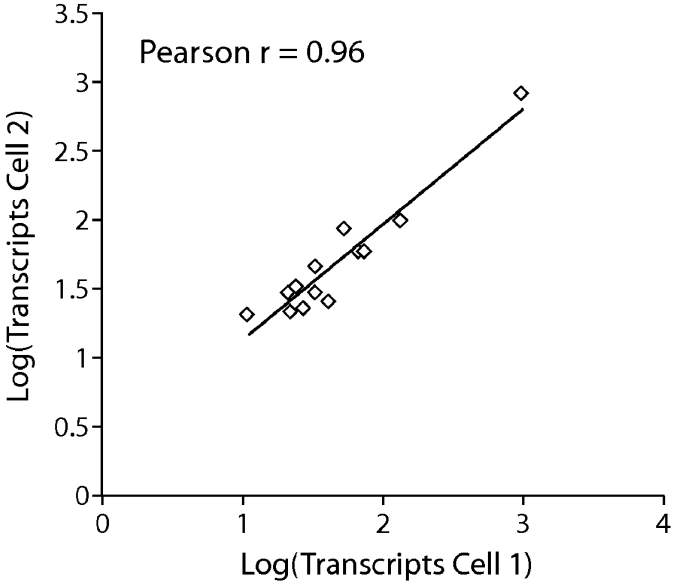

Pearson r = 0.96
Log(Transcripts Cell 2)
Log(Transcripts Cell 1)
0
0.5
1
1.5
2
2.5
3
3.5
0
1
2
3
4

Fig. 2G

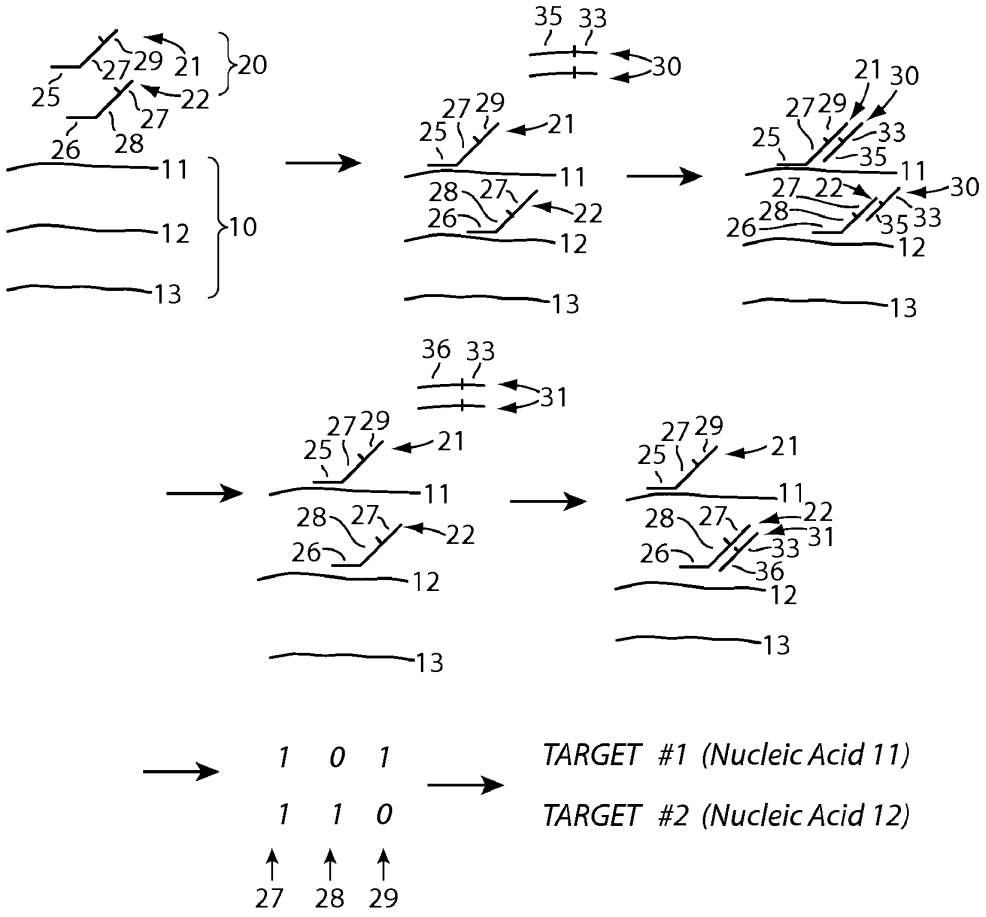

35 33
30
21 30
27 29
25
21
20
27 29
21
22
27
26 28
11
12
10
13
28 27
22
26
33
35
27 22
30
28
35 33
36 33
31
1 0 1
TARGET #1 (Nucleic Acid 11)
1 1 0
TARGET #2 (Nucleic Acid 12)
27 28 29

Fig. 3A

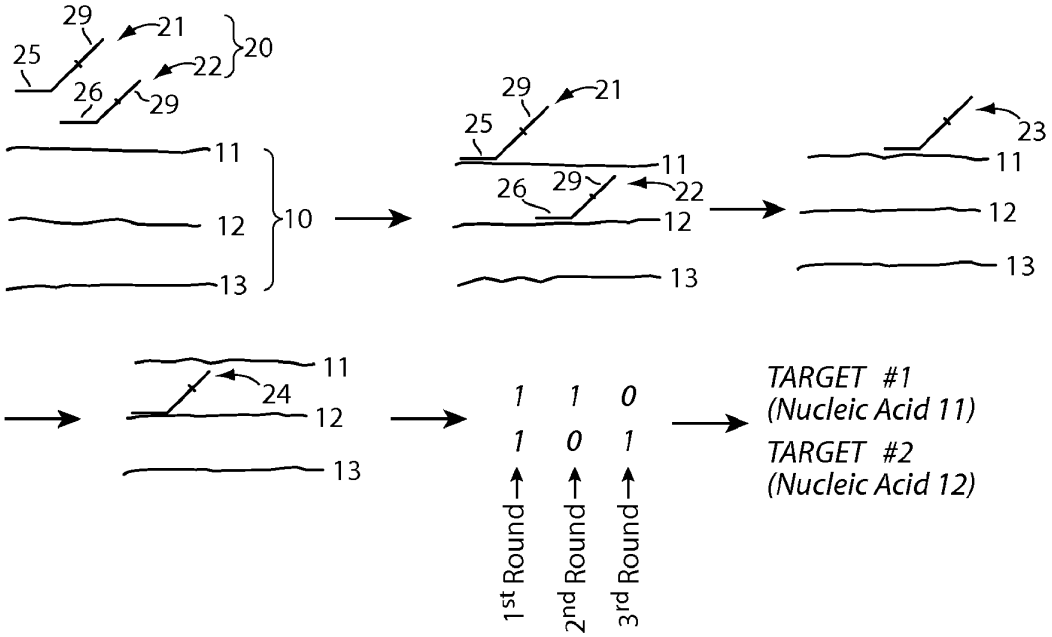


Fig. 3B

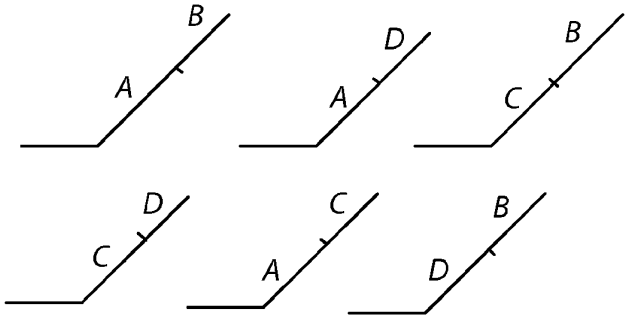


Fig. 4A

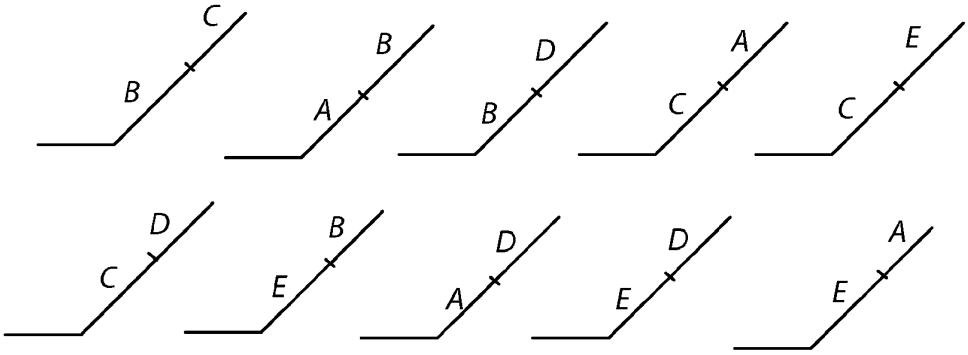


Fig. 4B

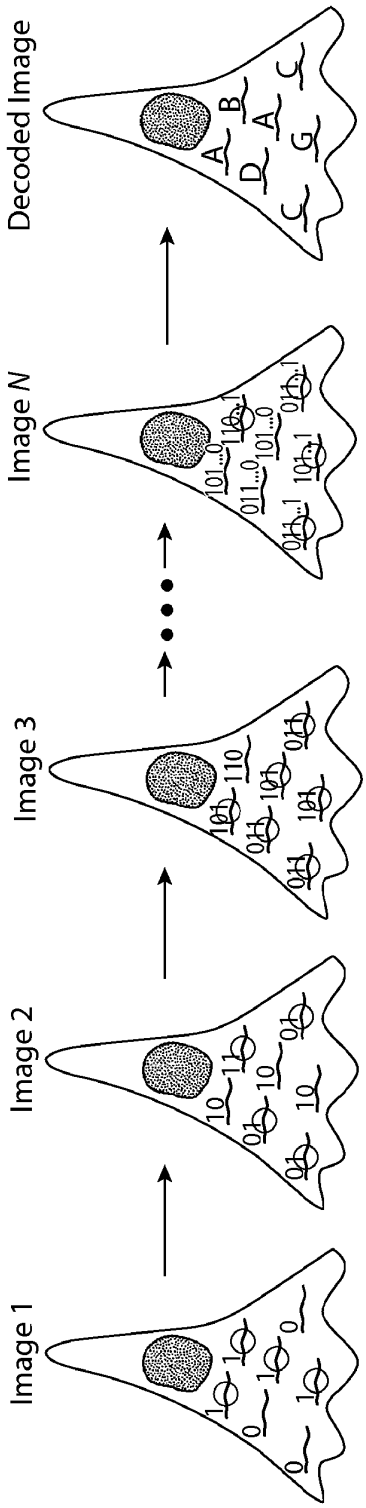


Fig. 5A

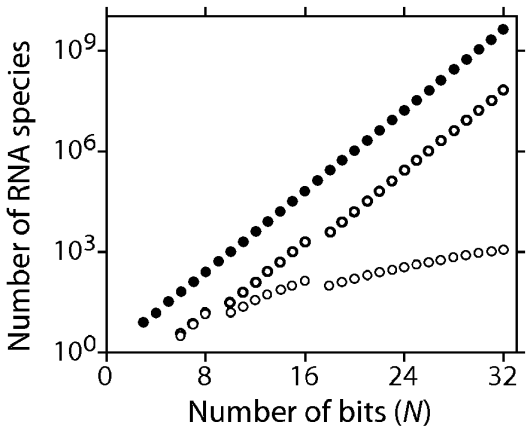


Fig. 5B

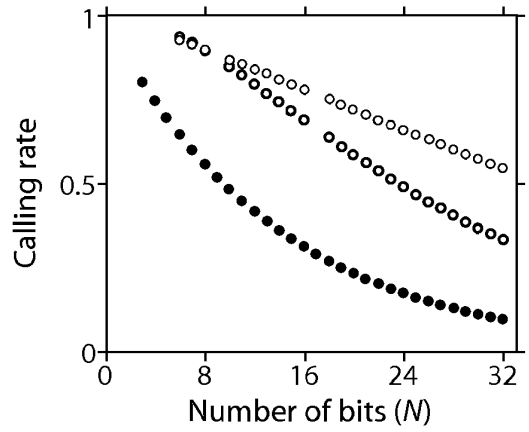


Fig. 5C

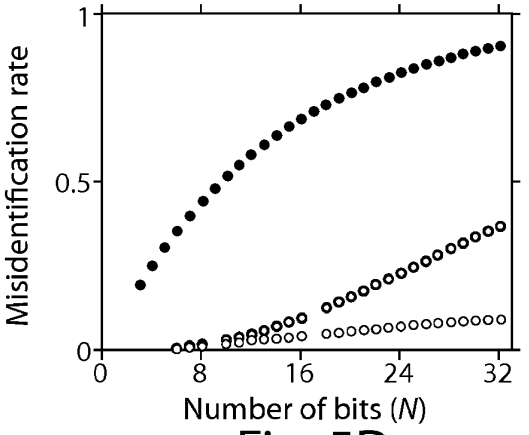


Fig. 5D

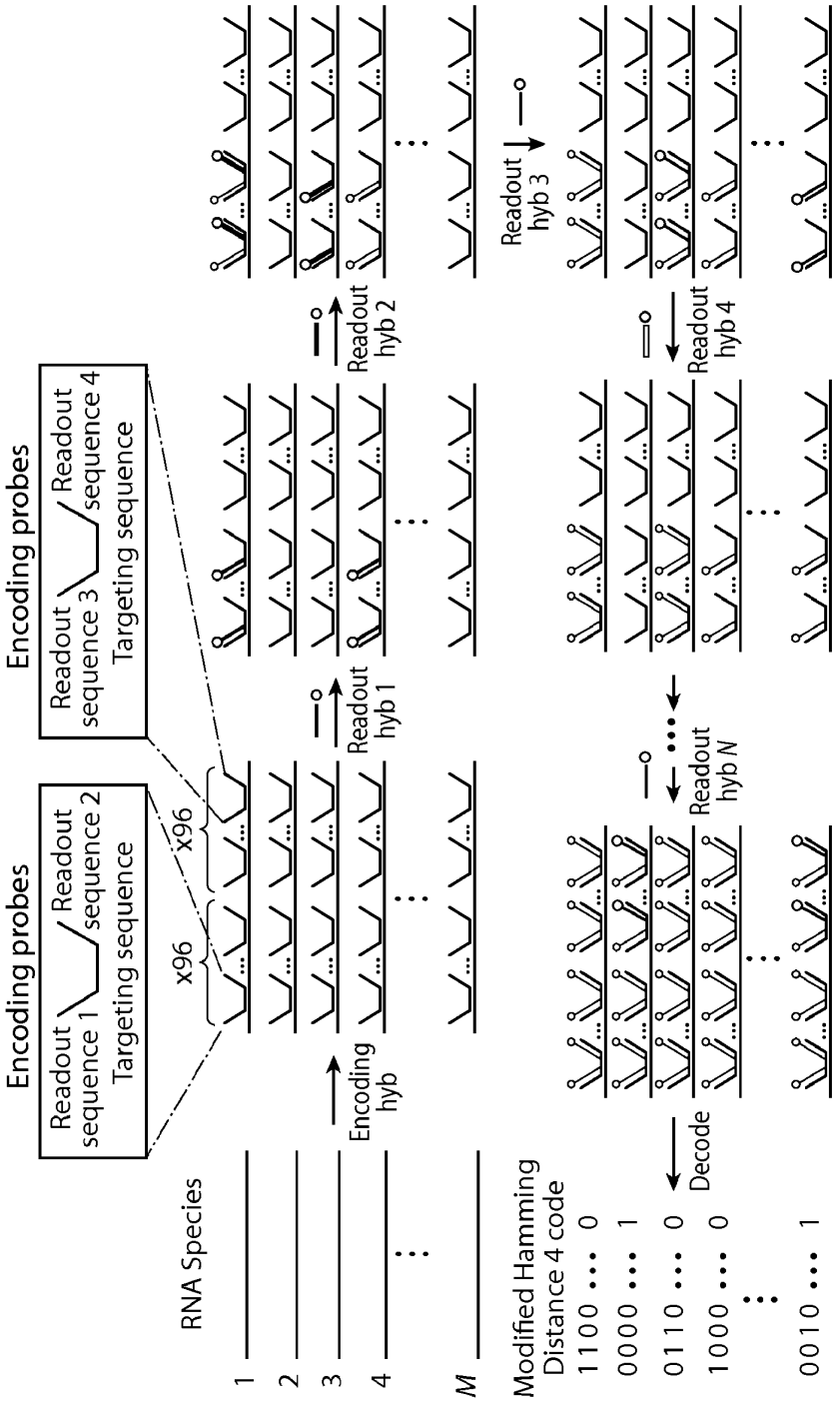

Encoding probes
Readout sequence 1
Readout sequence 2
Targeting sequence
Encoding probes
Readout sequence 3
Readout sequence 4
Targeting sequence
x96
x96
RNA Species
1
2
3
4
M
Encoding hyb
Readout hyb 1
Readout hyb 2
Readout hyb 3
Readout hyb 4
Readout hyb N
Modified Hamming Distance 4 code
1100 ... 0
0000 ... 1
0110 ... 0
1000 ... 0
0010 ... 1
Decode

Fig. 5E

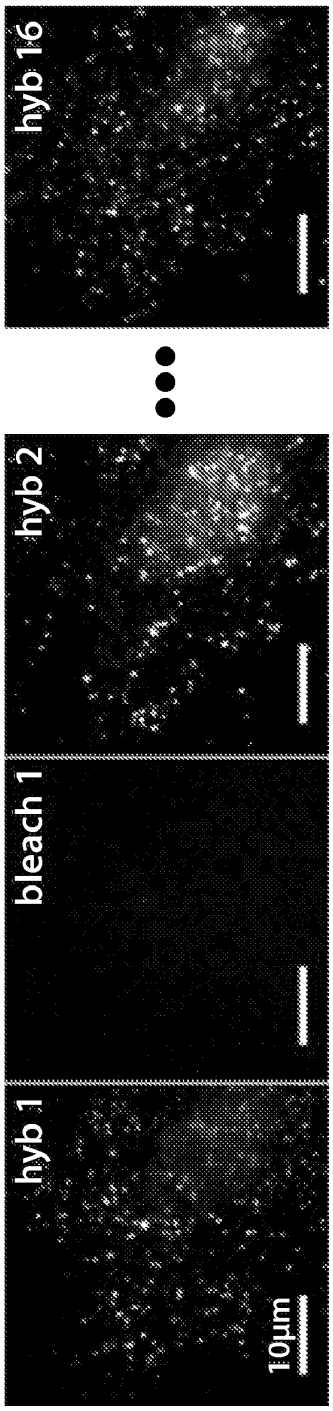


Fig. 6A

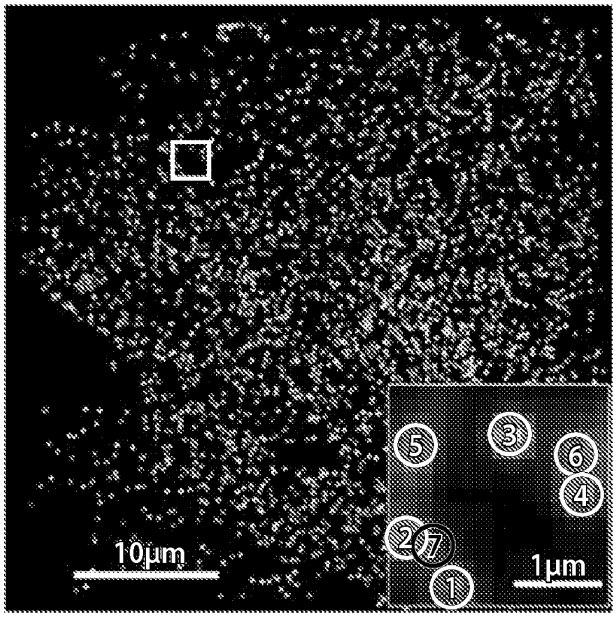

10µm
1µm
1
2
3
4
5
6
7

Fig. 6B

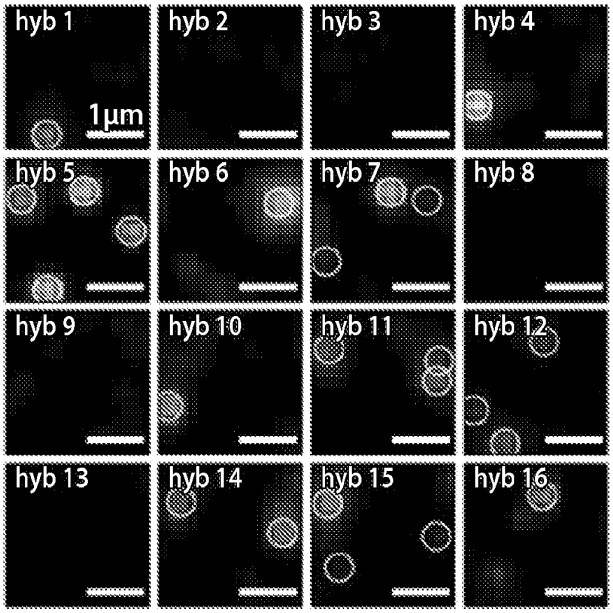

hyb 1
1µm
hyb 2
hyb 3
hyb 4
hyb 5
hyb 6
hyb 7
hyb 8
hyb 9
hyb 10
hyb 11
hyb 12
hyb 13
hyb 14
hyb 15
hyb 16

Fig. 6C

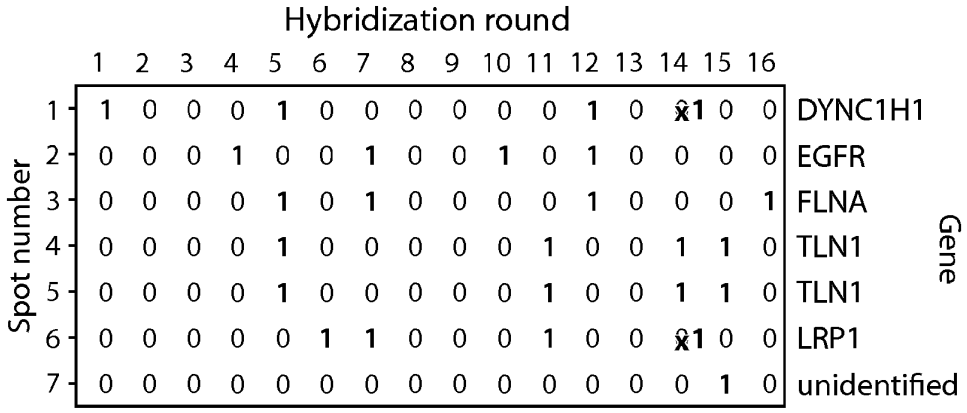

Hybridization round
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Spot number
1 1 0 0 0 1 0 0 0 0 0 0 1 0 1 0 0 DYNC1H1
2 0 0 0 1 0 0 1 0 0 1 0 1 0 0 0 0 EGFR
3 0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 1 FLNA
4 0 0 0 0 1 0 0 0 0 0 1 0 0 1 1 0 TLN1
5 0 0 0 0 1 0 0 0 0 0 1 0 0 1 1 0 TLN1
6 0 0 0 0 0 1 1 0 0 0 1 0 0 1 0 0 LRP1
7 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 unidentified
Gene

Fig. 6D

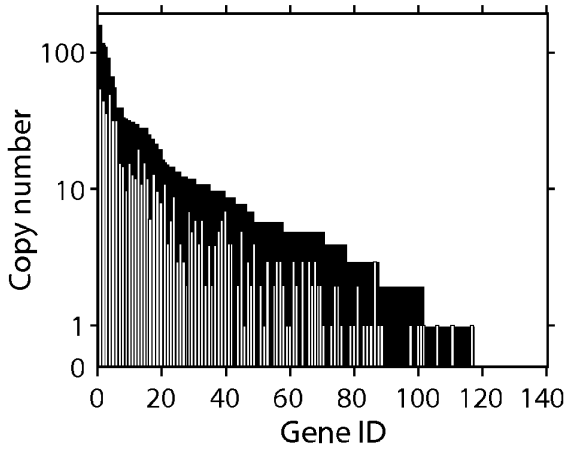

Copy number
100
10
1
0
0 20 40 60 80 100 120 140
Gene ID

Fig. 6E

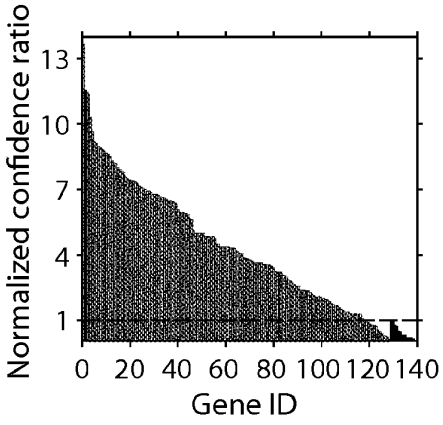


Fig. 6F

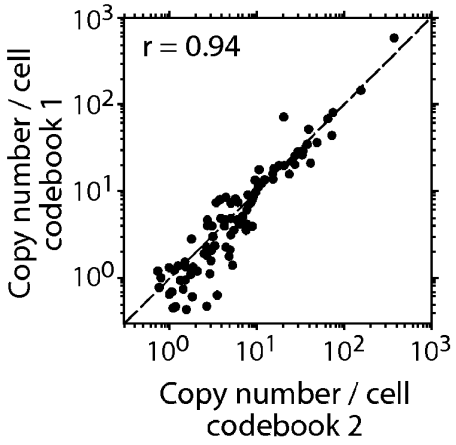


Fig. 6G

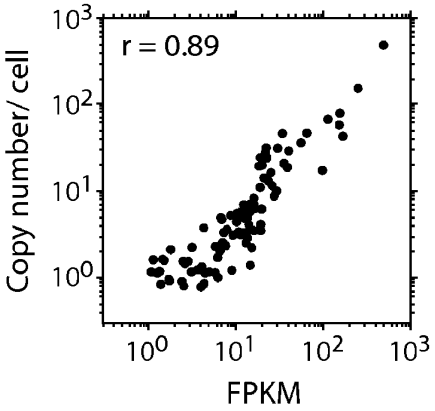


Fig. 6H

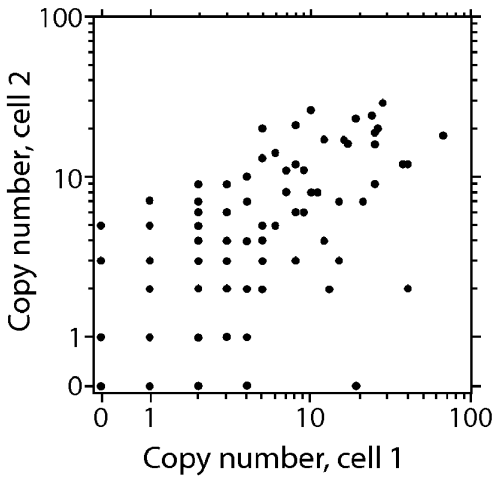

100
10
1
0
0
1
10
100
Copy number, cell 2
Copy number, cell 1

Fig. 7A

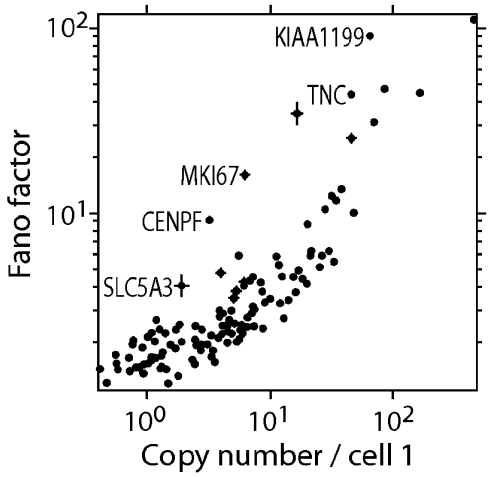

10^2
10^1
10^0
10^1
10^2
KIAA1199
TNC
MKI67
CENPF
SLC5A3
Fano factor
Copy number / cell 1

Fig. 7B

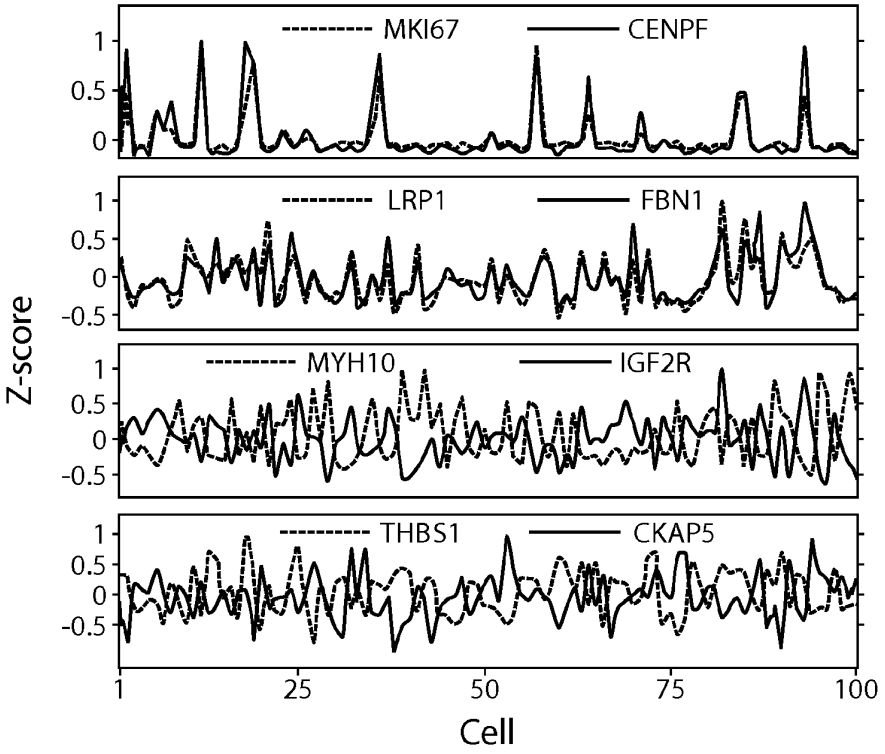

MKI67
CENPF
LRP1
FBN1
MYH10
IGF2R
THBS1
CKAP5
Z-score
1
0.5
0
-0.5
1
25
50
75
100
Cell

Fig. 7C

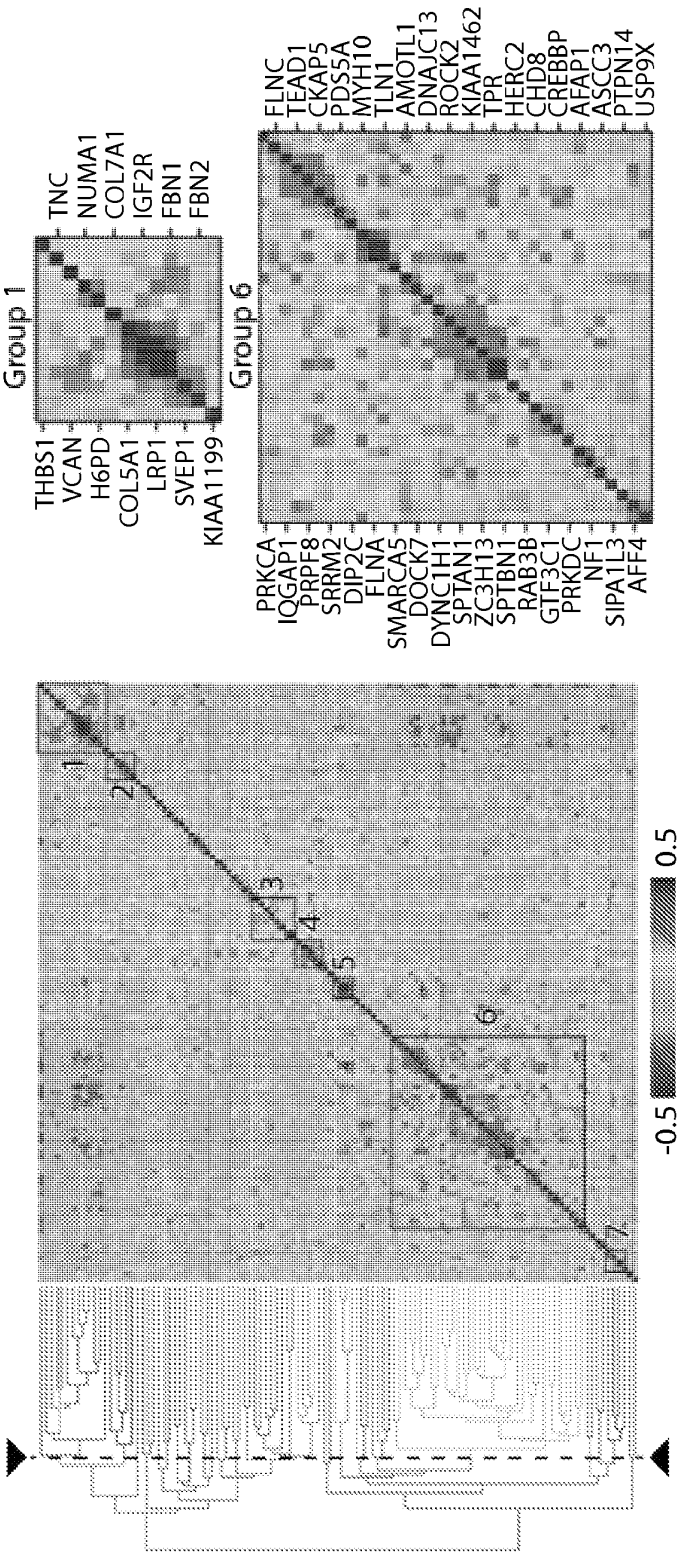

Group 1
TNC
NUMA1
COL7A1
IGF2R
FBN1
FBN2
THBS1
VCAN
H6PD
COL5A1
LRP1
SVEP1
KIAA1199
Group 6
FLNC
TEAD1
CKAP5
PDS5A
MYH10
TLN1
AMOTL1
DNAJC13
ROCK2
KIAA1462
TPR
HERC2
CHD8
CREBBP
AFAP1
ASCC3
PTPN14
USP9X
PRKCA
IQGAP1
PRPF8
SRRM2
DIP2C
FLNA
SMARCA5
DOCK7
DYNC1H1
SPTAN1
ZC3H13
SPTBN1
RAB3B
GTF3C1
PRKDC
NF1
SIPA1L3
AFF4
1
2
3
4
5
6
7
0.5
-0.5

Fig. 7D

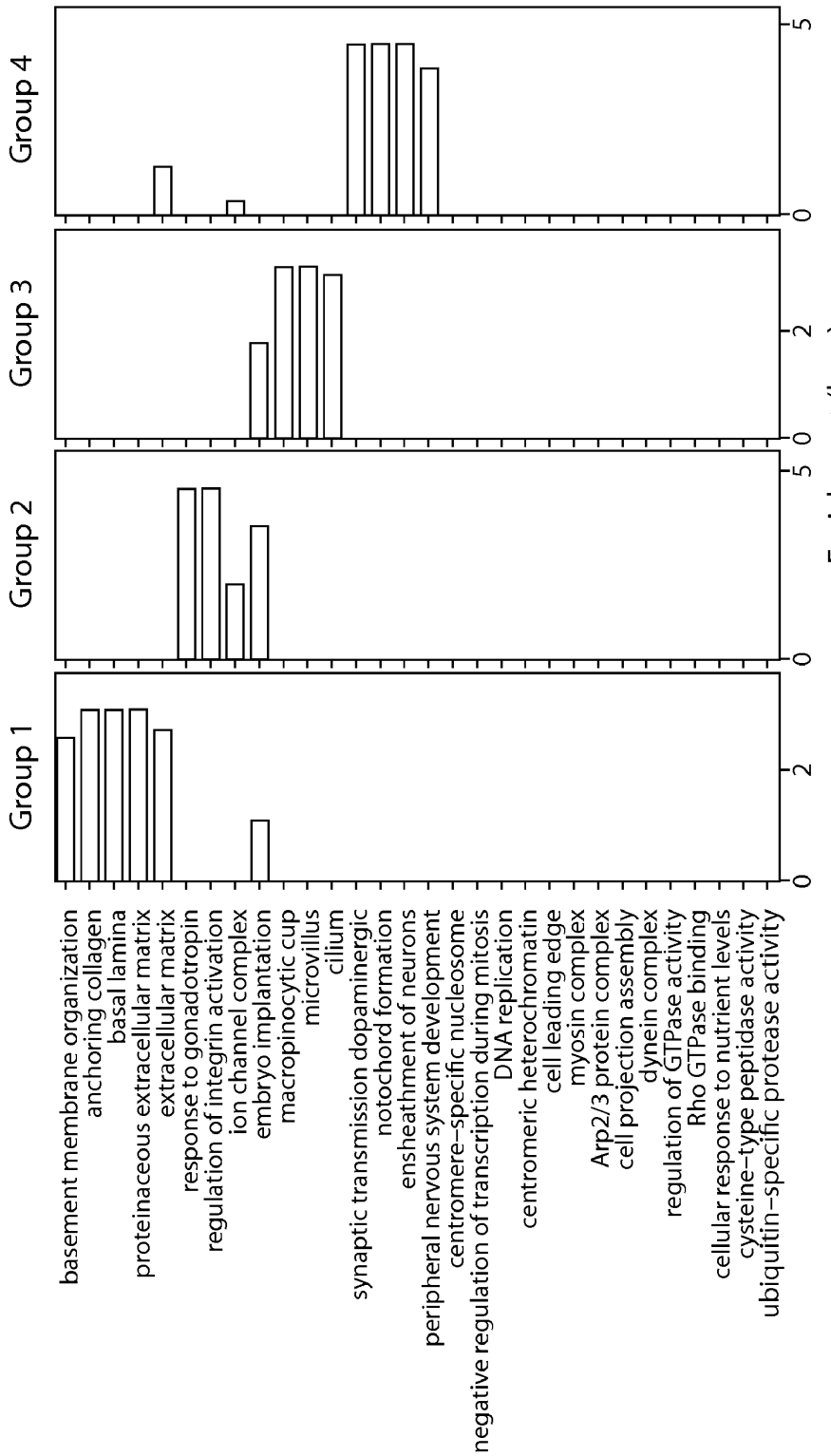

Group 1
Group 2
Group 3
Group 4
0
2
0
5
0
2
0
5
Enrichment (log2)
basement membrane organization
anchoring collagen
basal lamina
proteinaceous extracellular matrix
extracellular matrix
response to gonadotropin
regulation of integrin activation
ion channel complex
embryo implantation
macropinocytic cup
microvillus
cilium
synaptic transmission dopaminergic
notochord formation
ensheathment of neurons
peripheral nervous system development
centromere−specific nucleosome
negative regulation of transcription during mitosis
DNA replication
centromeric heterochromatin
cell leading edge
myosin complex
Arp2/3 protein complex
cell projection assembly
dynein complex
regulation of GTPase activity
Rho GTPase binding
cellular response to nutrient levels
cysteine−type peptidase activity
ubiquitin−specific protease activity

Fig. 7E

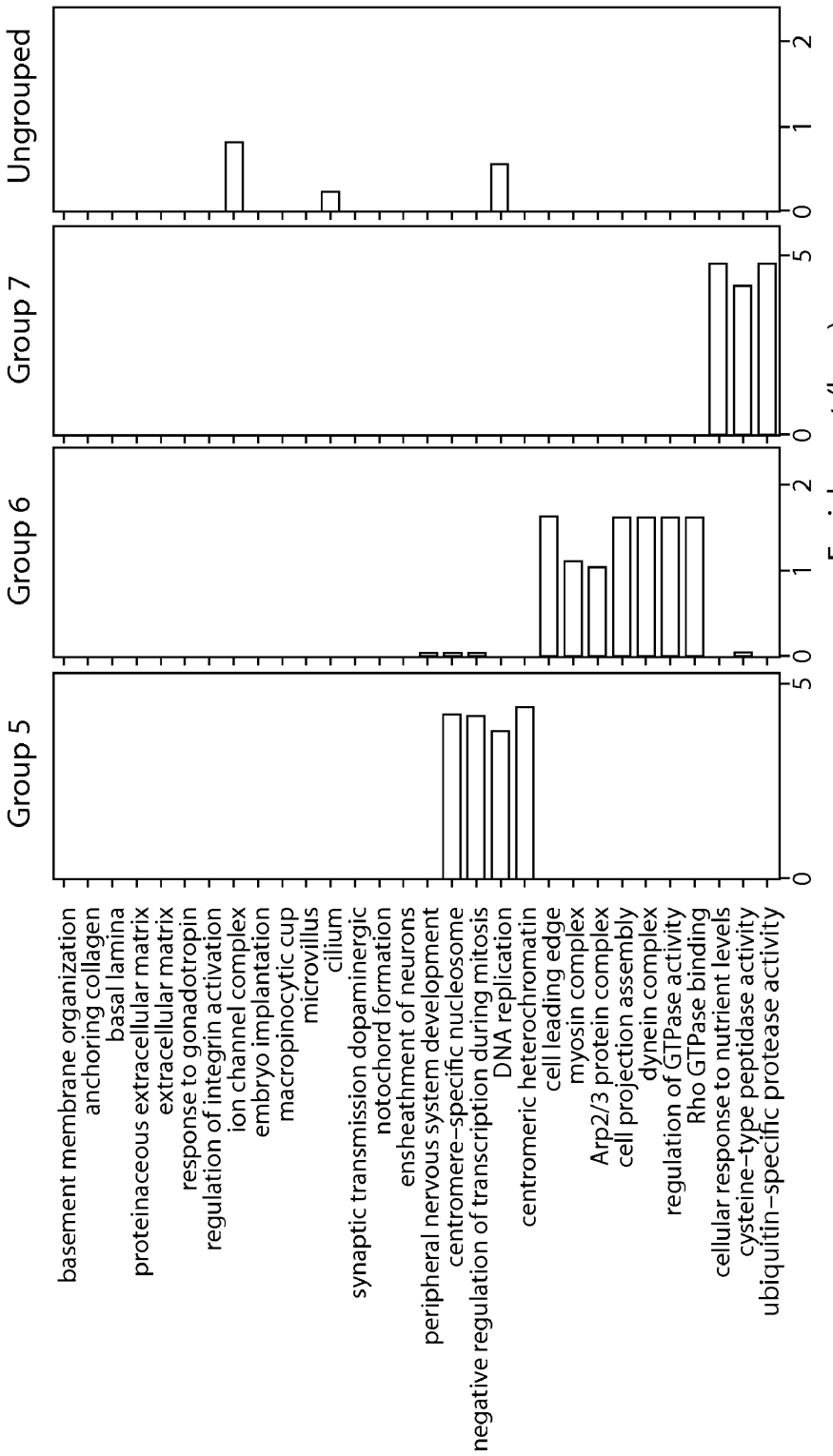


Fig. 7E Continued

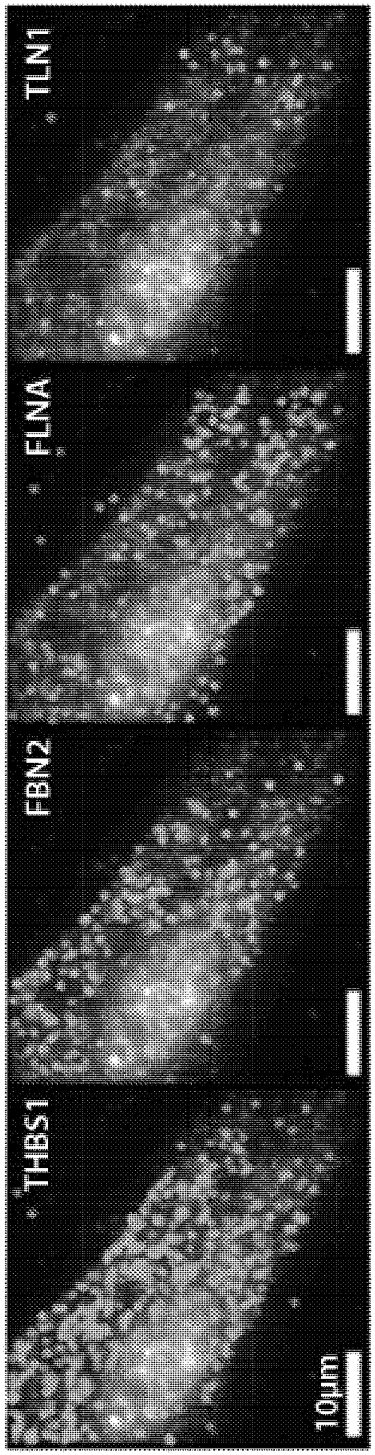


Fig. 8A

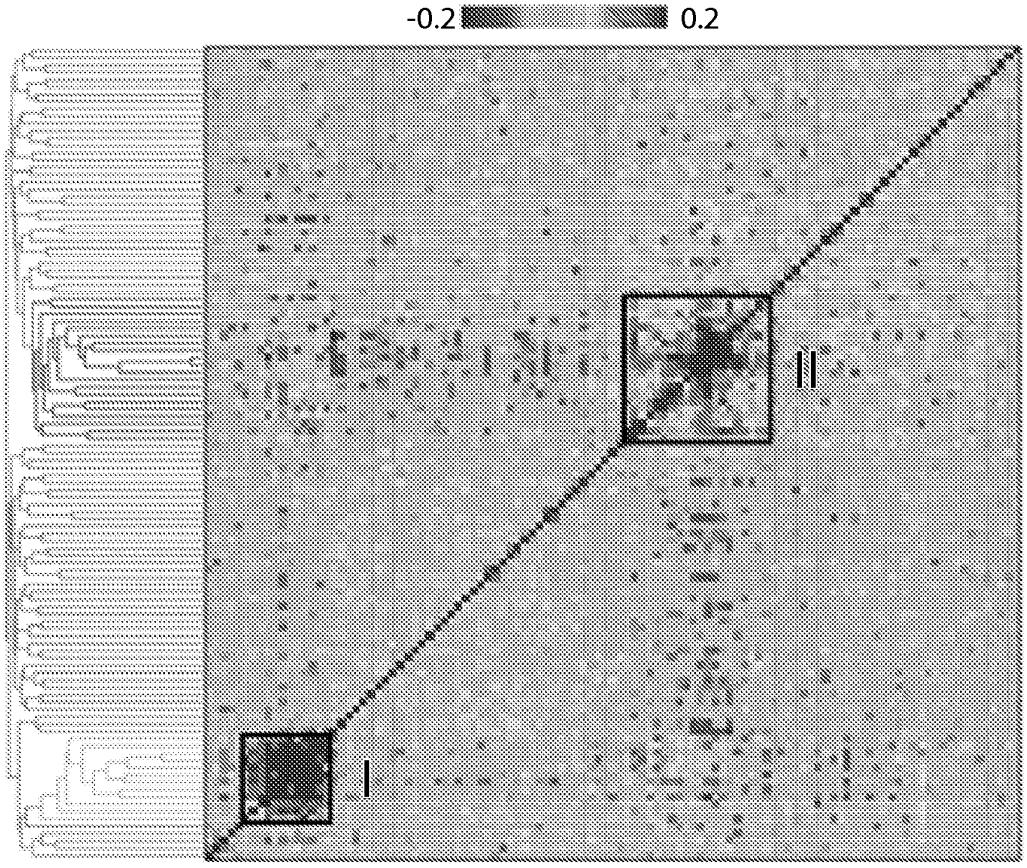


Fig. 8B

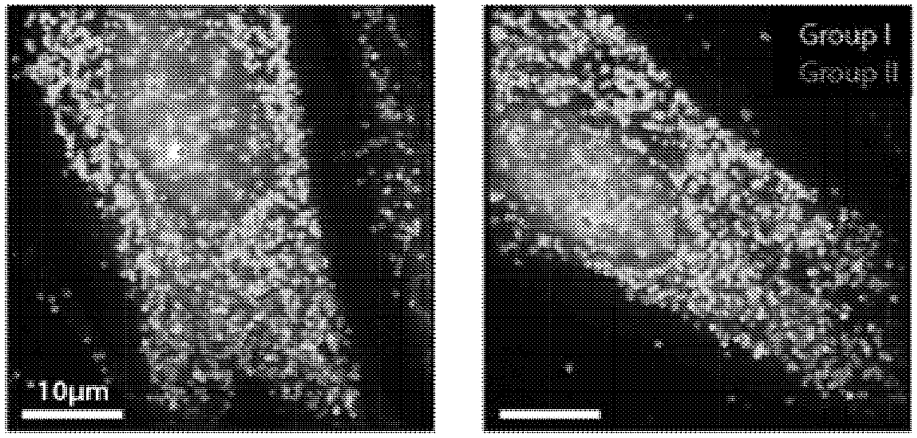


Fig. 8C

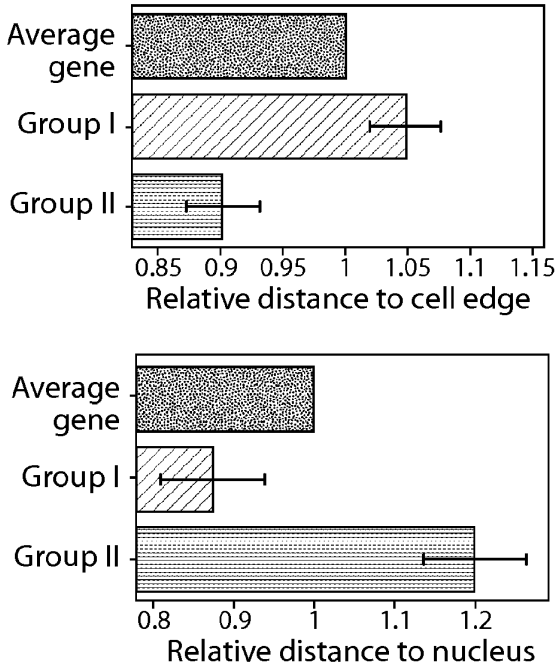

Average gene
Group I
Group II
0.85 0.9 0.95 1 1.05 1.1 1.15
Relative distance to cell edge
Average gene
Group I
Group II
0.8 0.9 1 1.1 1.2
Relative distance to nucleus

Fig. 8D

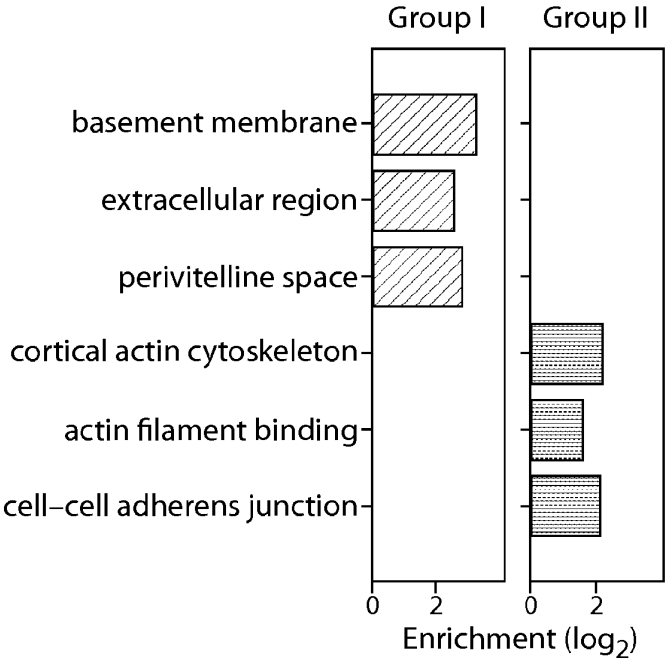

Group I
Group II
basement membrane
extracellular region
perivitelline space
cortical actin cytoskeleton
actin filament binding
cell–cell adherens junction
0 2
0 2
Enrichment (log2)

Fig. 8E

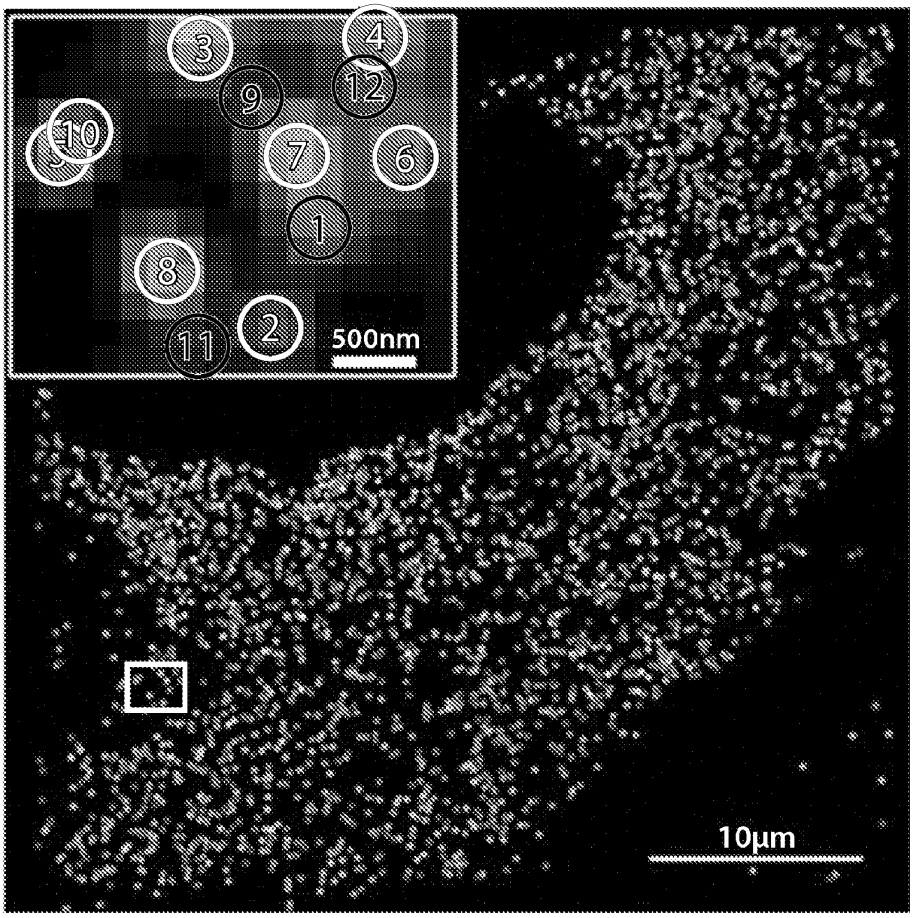


Fig. 9A

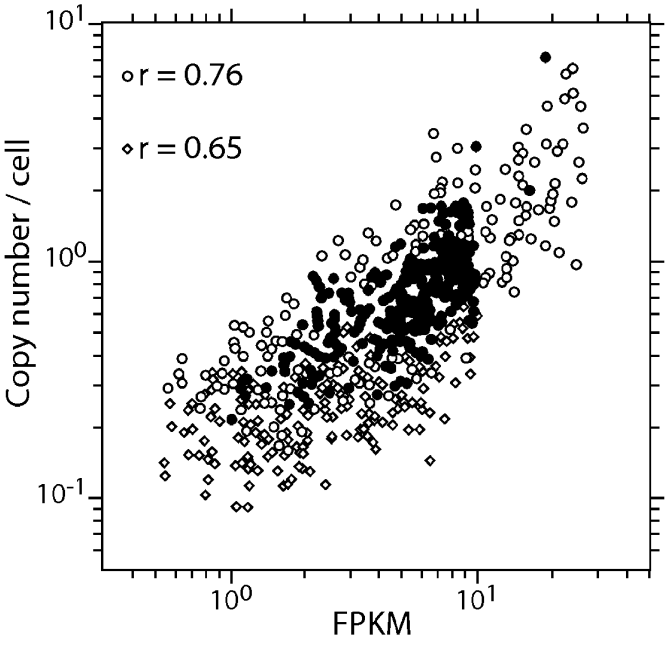

r = 0.76
r = 0.65
Copy number / cell
FPKM

Fig. 9B

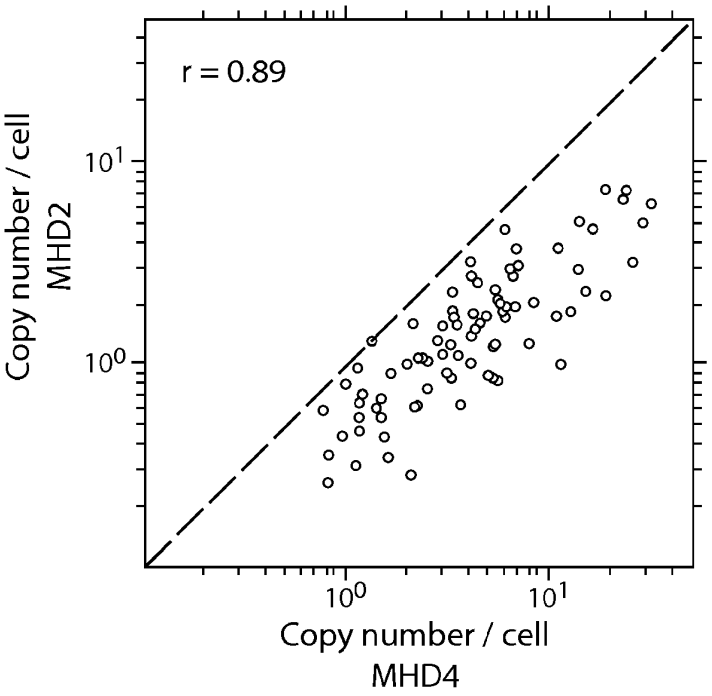

r = 0.89
Copy number / cell
MHD2
Copy number / cell
MHD4

Fig. 9C

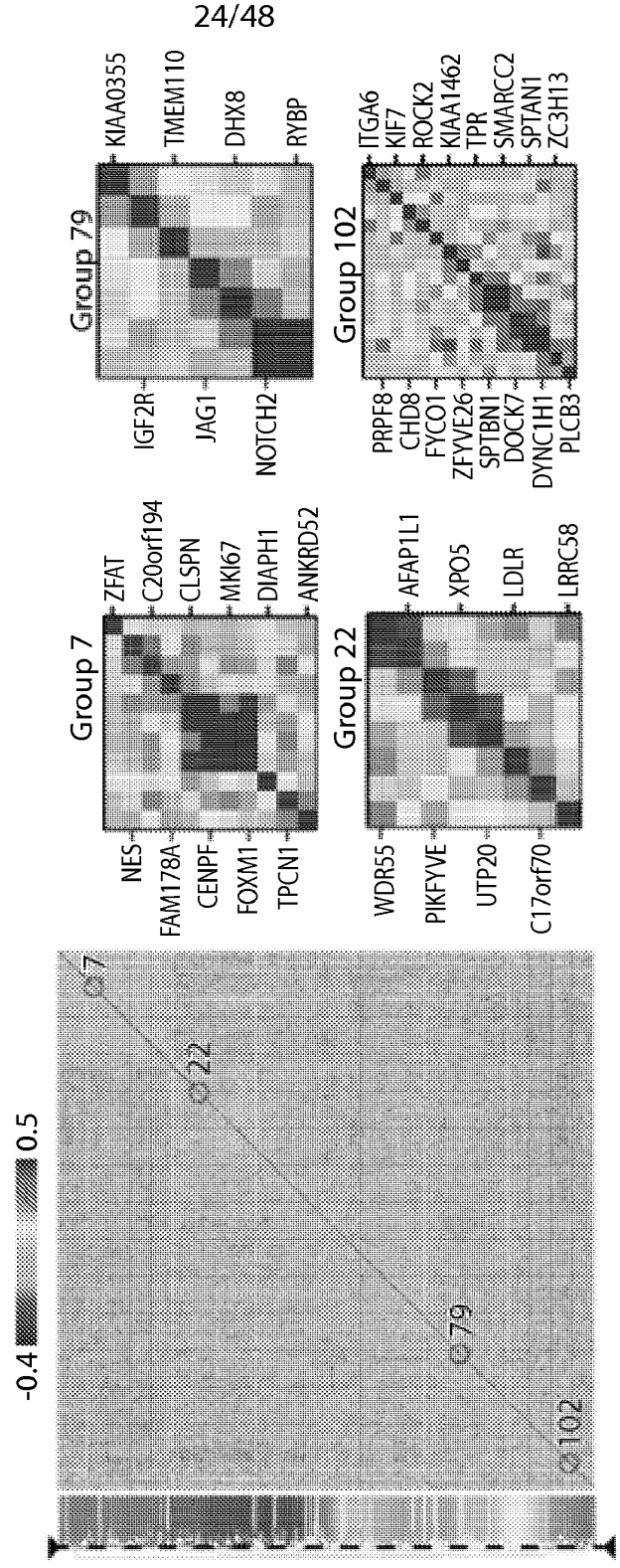

-0.4
0.5
g7
g22
g79
g102
Group 7
NES
FAM178A
CENPF
FOXM1
TPCN1
ZFAT
C20orf194
CLSPN
MKI67
DIAPH1
ANKRD52
Group 22
WDR55
PIKFYVE
UTP20
C17orf70
AFAP1L1
XPO5
LDLR
LRRC58
Group 79
IGF2R
JAG1
NOTCH2
KIAA0355
TMEM110
DHX8
RYBP
Group 102
PRPF8
CHD8
FYCO1
ZFYVE26
SPTBN1
DOCK7
DYNC1H1
PLCB3
ITGA6
KIF7
ROCK2
KIAA1462
TPR
SMARCC2
SPTAN1
ZC3H13

Fig. 10A

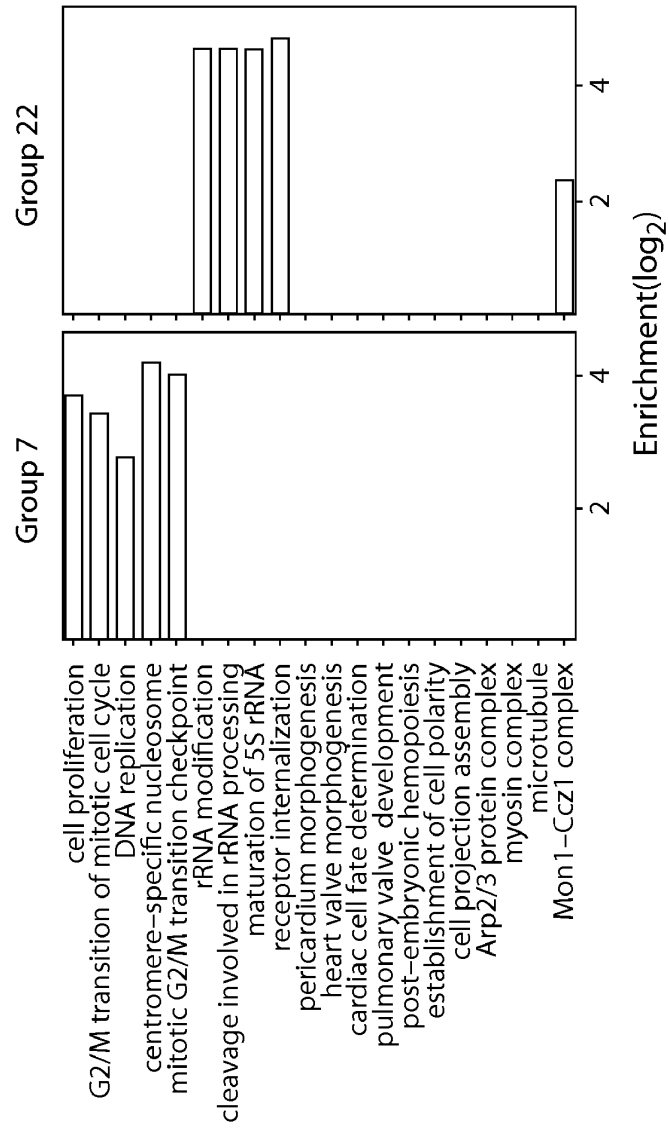

Group 7
Group 22
cell proliferation
G2/M transition of mitotic cell cycle
DNA replication
centromere−specific nucleosome
mitotic G2/M transition checkpoint
rRNA modification
cleavage involved in rRNA processing
maturation of 5S rRNA
receptor internalization
pericardium morphogenesis
heart valve morphogenesis
cardiac cell fate determination
pulmonary valve development
post−embryonic hemopoiesis
establishment of cell polarity
cell projection assembly
Arp2/3 protein complex
myosin complex
microtubule
Mon1−Ccz1 complex
2
4
2
4
Enrichment(log2)

Fig. 10B

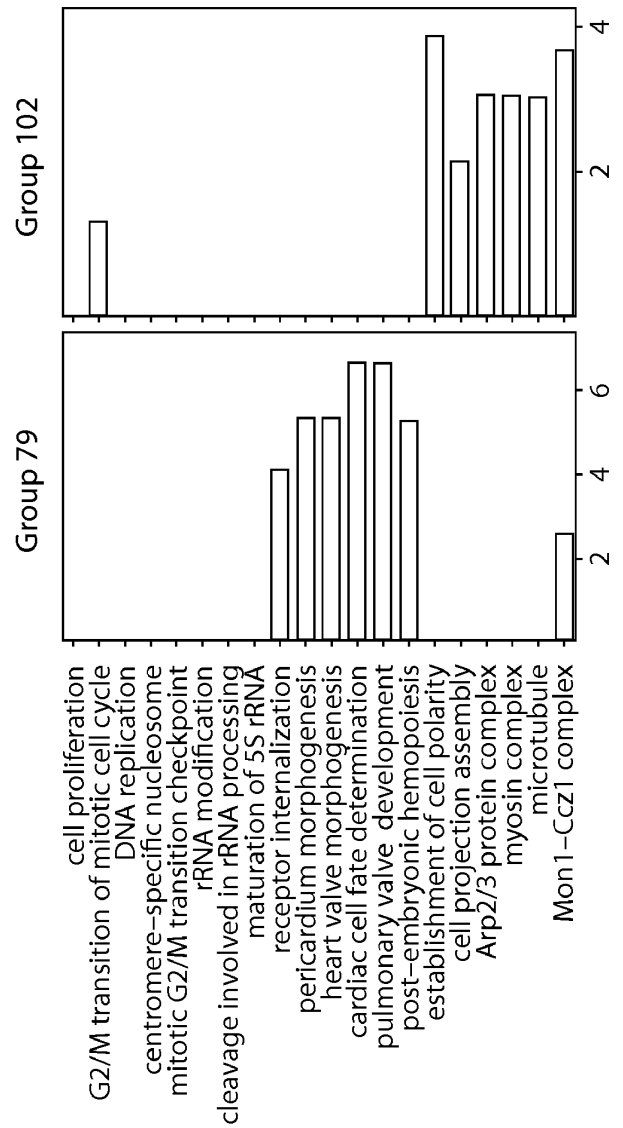


Fig. 10B Continued

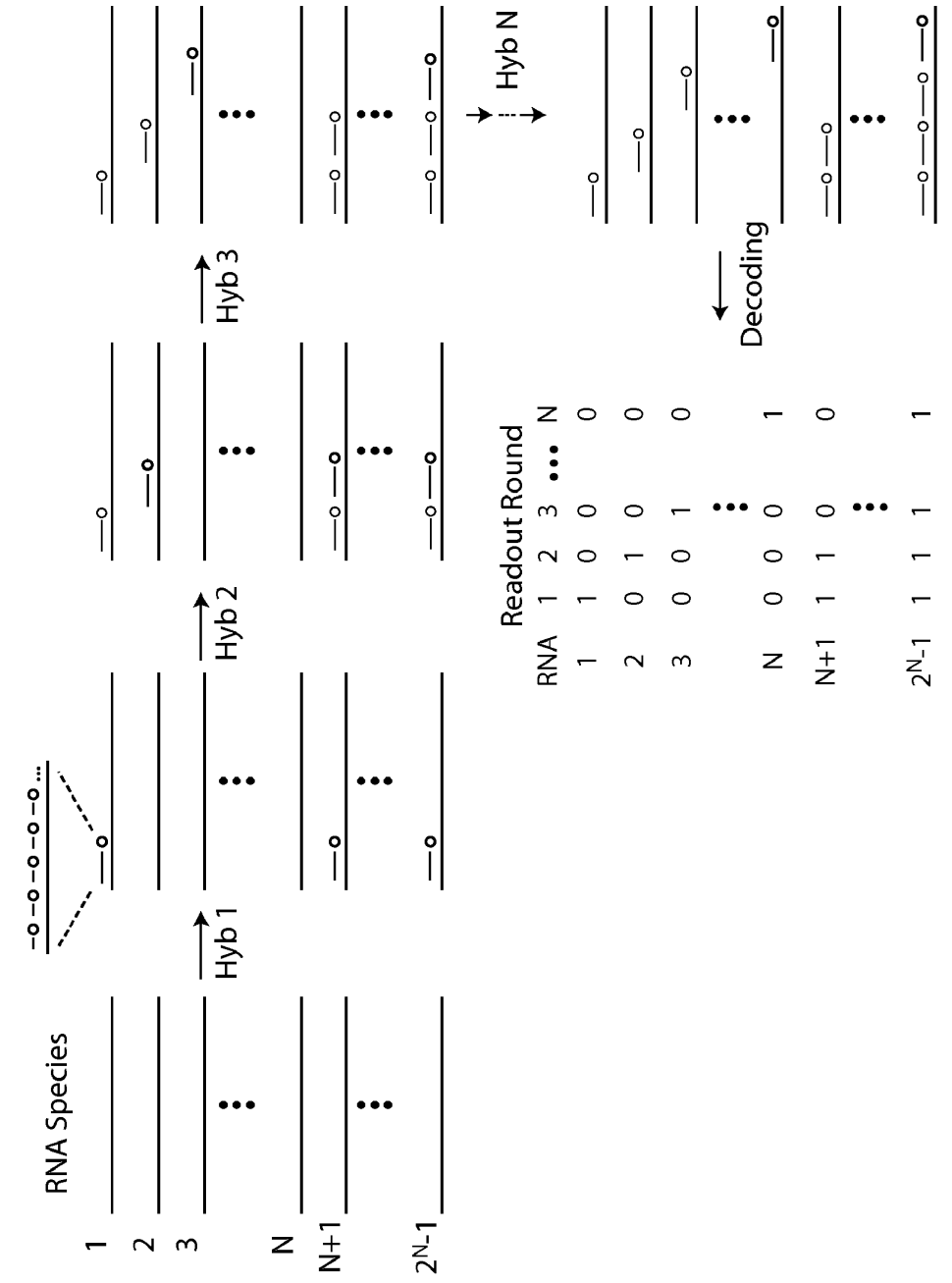

RNA Species
1
2
3
N
N+1
2N-1
Hyb 1
Hyb 2
Hyb 3
Hyb N
Readout Round
RNA 1 2 3 ... N
1 1 0 0 0
2 0 1 0 0
3 0 0 1 0
N 0 0 0 1
N+1 1 1 0 0
2N-1 1 1 1 1
Decoding

Fig. 11

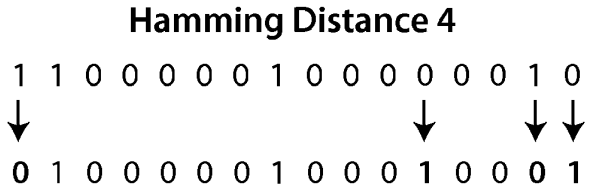


Fig. 12A

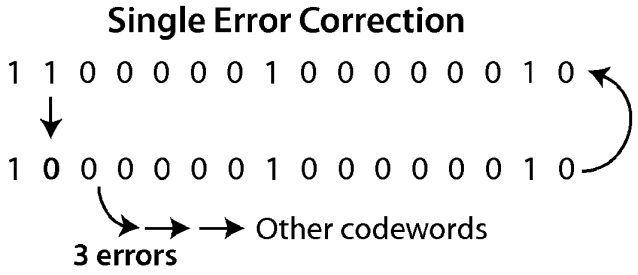


Fig. 12B

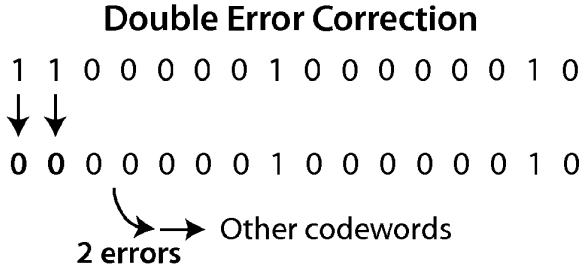


Fig. 12C

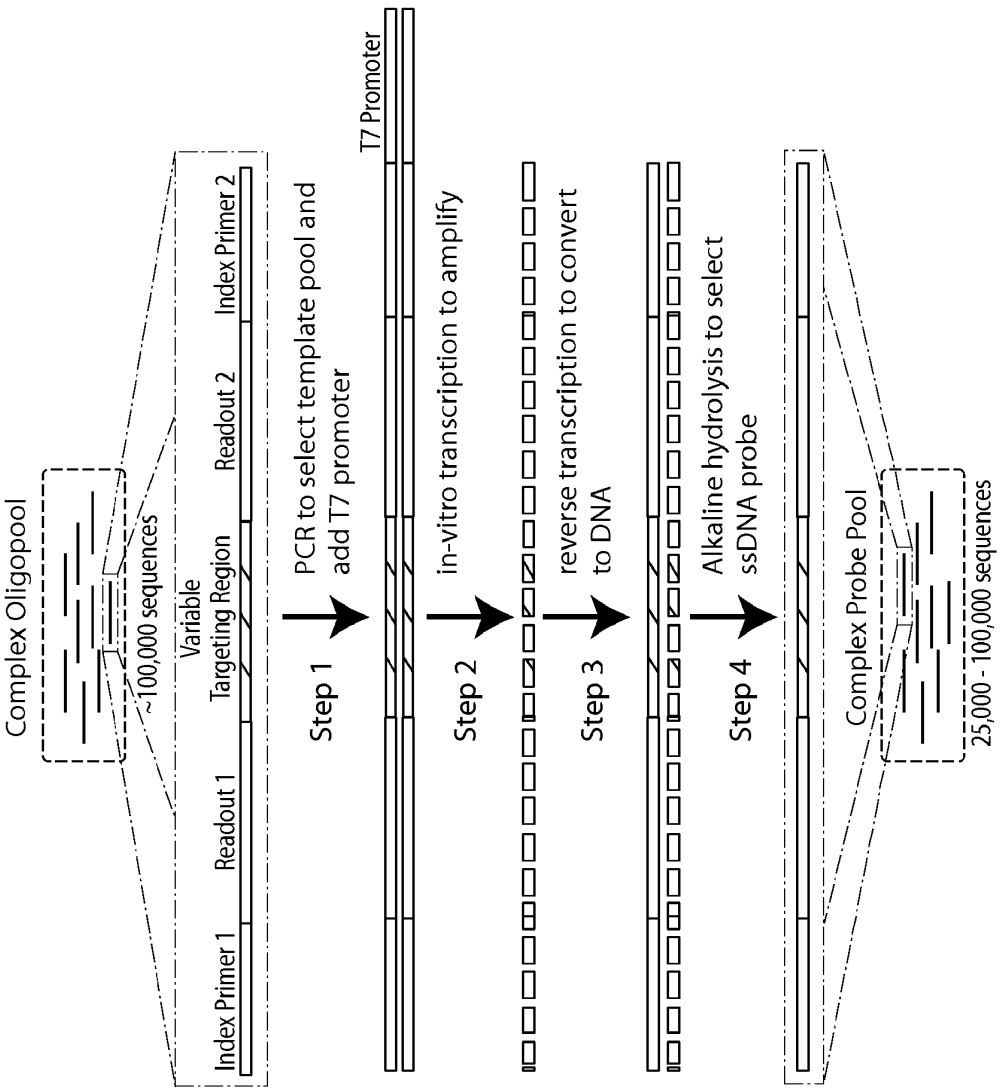

Complex Oligopool
~100,000 sequences
Index Primer 1
Readout 1
Variable
Targeting Region
Readout 2
Index Primer 2
Step 1
PCR to select template pool and
add T7 promoter
T7 Promoter
Step 2
in-vitro transcription to amplify
Step 3
reverse transcription to convert
to DNA
Step 4
Alkaline hydrolysis to select
ssDNA probe
Complex Probe Pool
25,000 - 100,000 sequences

Fig. 13

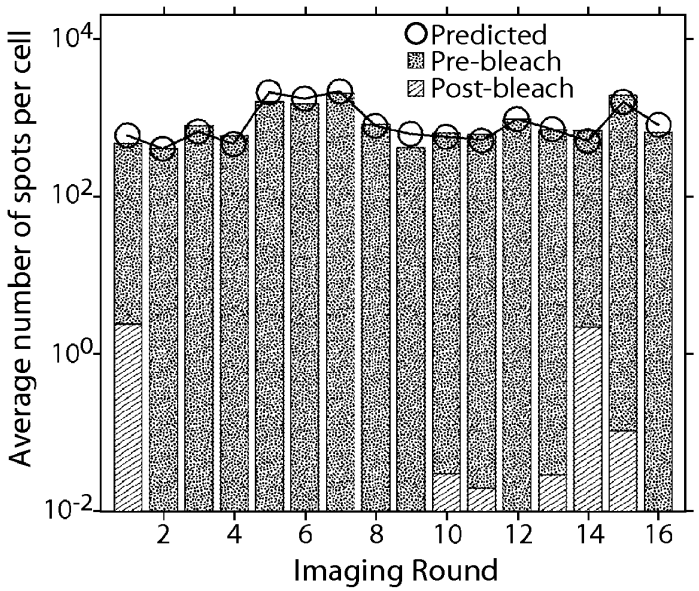


Fig. 14A

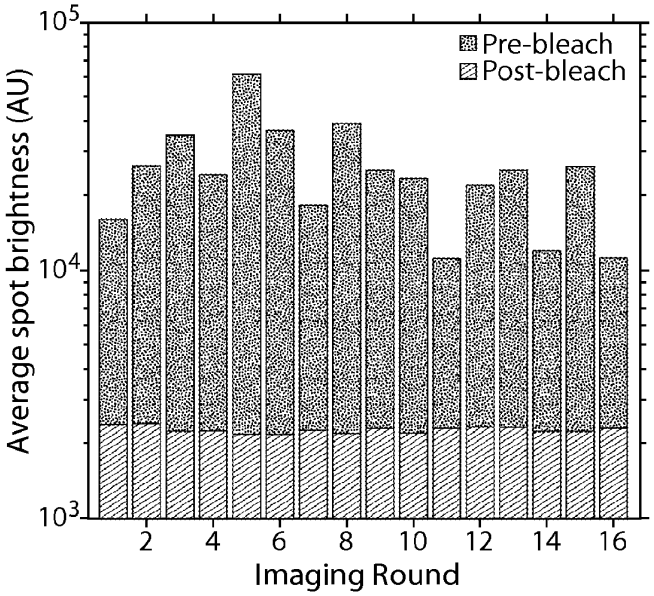


Fig. 14B

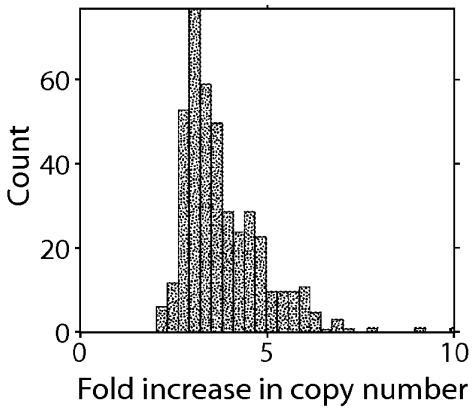


Fig. 15A

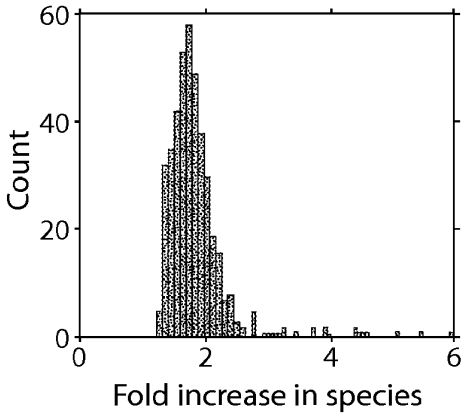


Fig. 15B

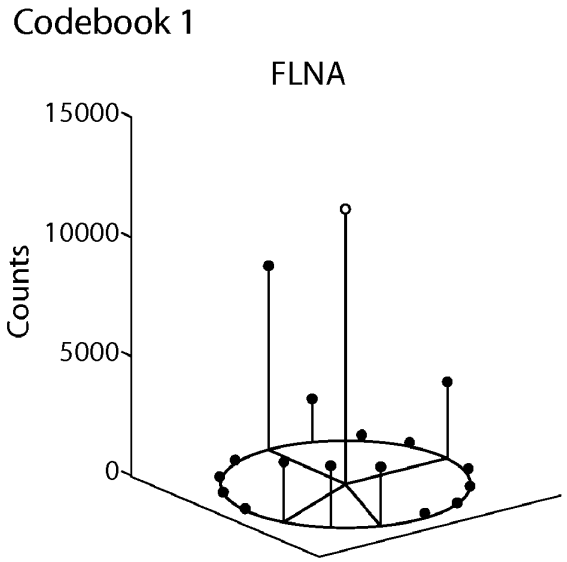


Fig. 16A

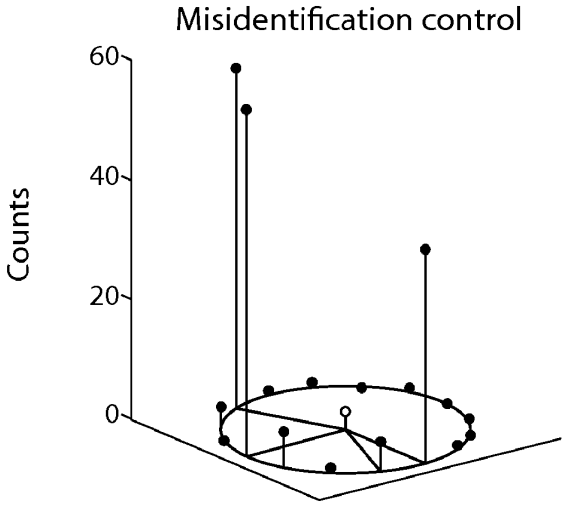


Fig. 16B

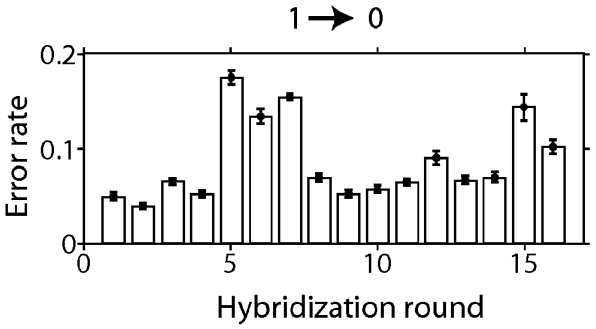


Fig. 16C

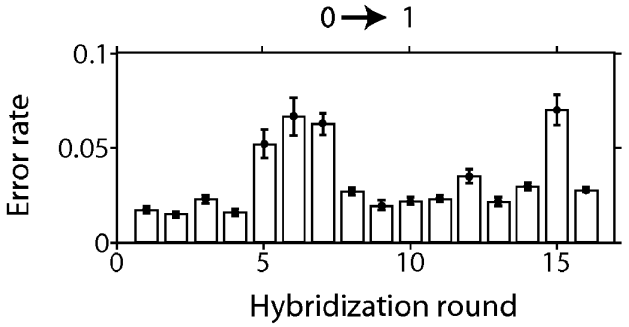


Fig. 16D

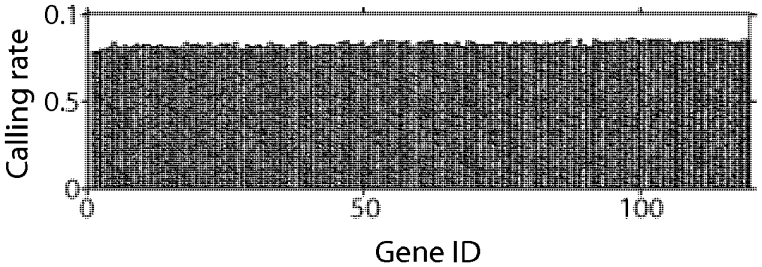


Fig. 16E

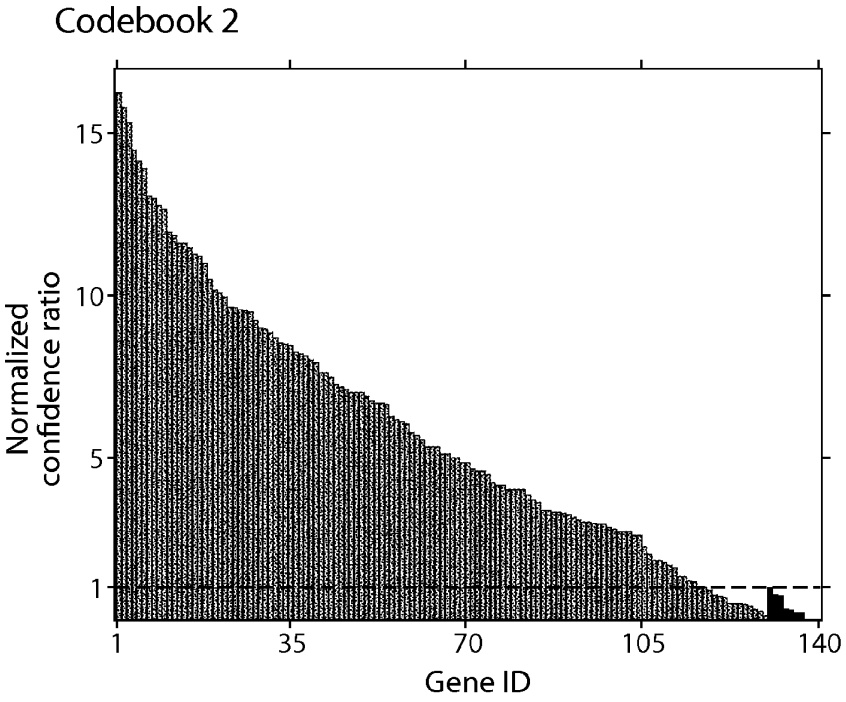


Fig. 17A

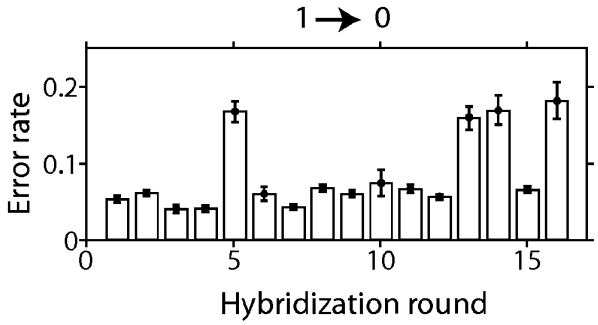


Fig. 17B

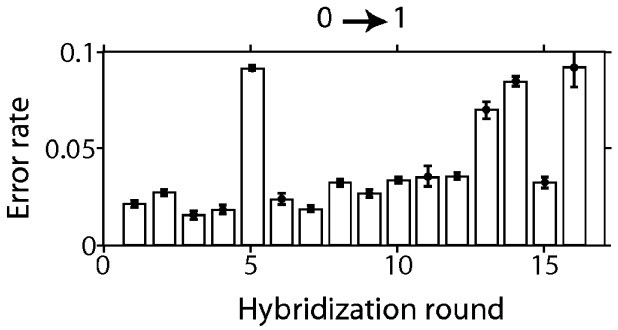


Fig. 17C

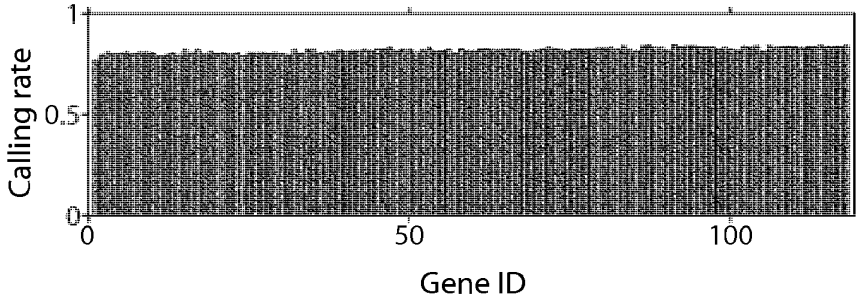


Fig. 17D

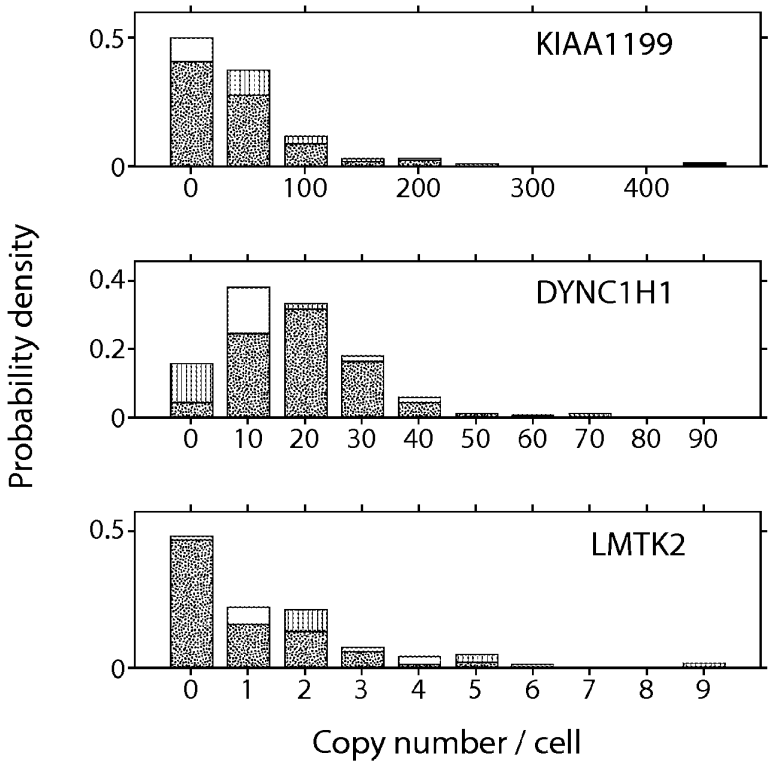

KIAA1199
DYNC1H1
LMTK2
Probability density
Copy number / cell
0.5
0
0.4
0.2
0 100 200 300 400
0 10 20 30 40 50 60 70 80 90
0 1 2 3 4 5 6 7 8 9

Fig. 18A

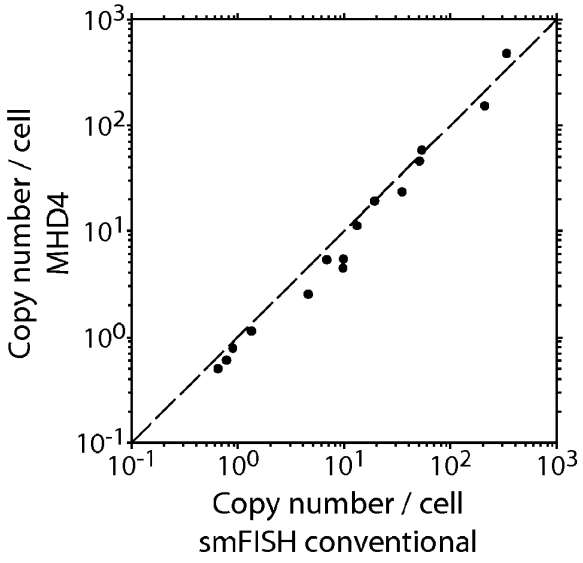


Fig. 18B

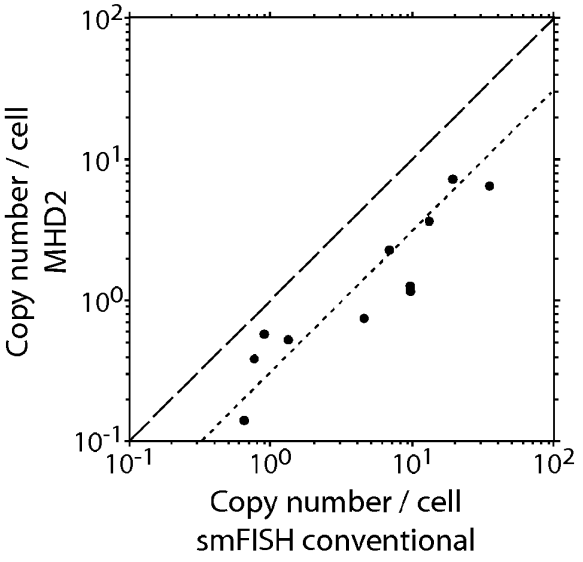


Fig. 18C

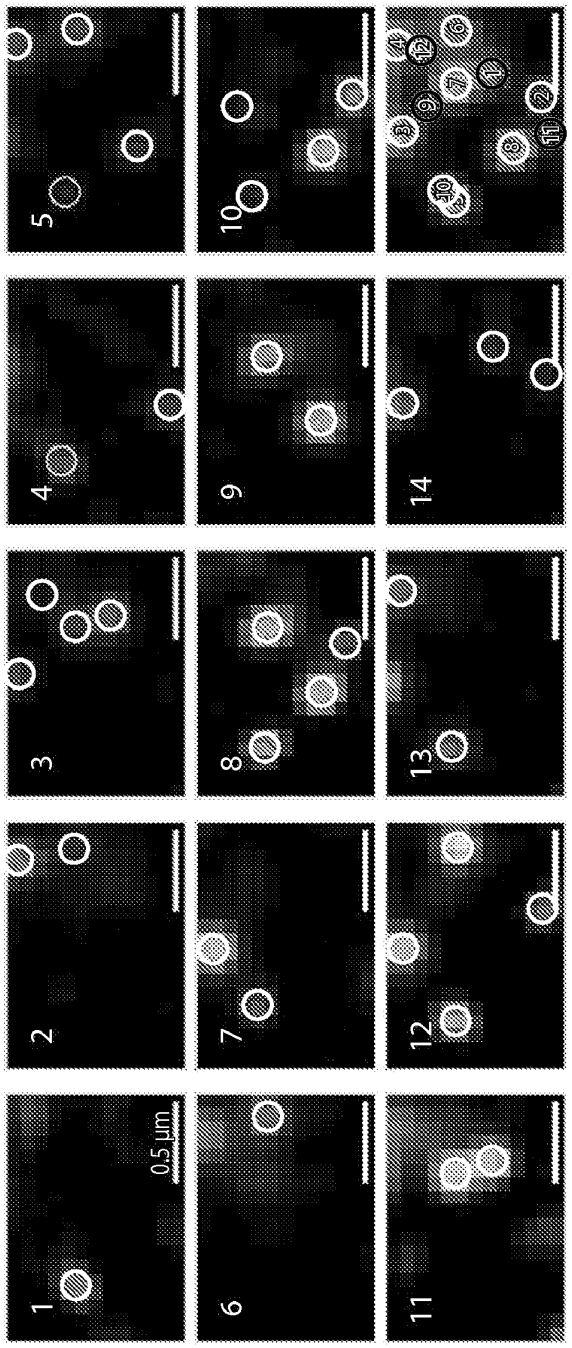


Fig. 19A

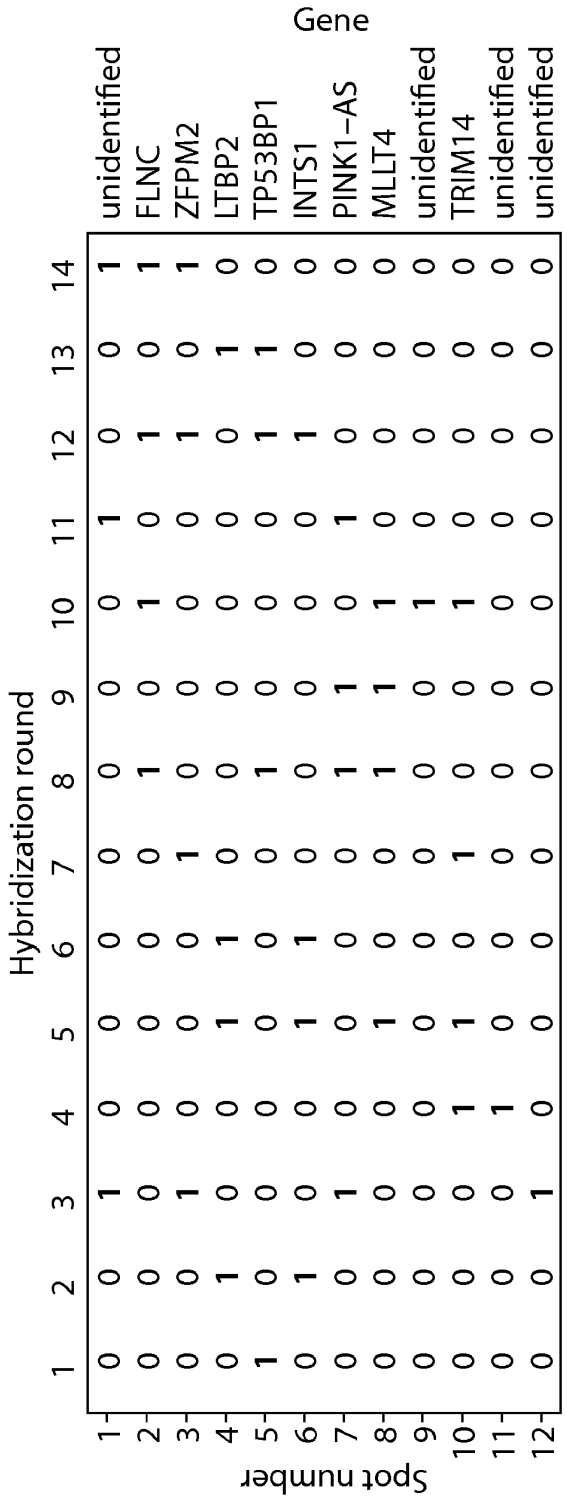

Gene
unidentified
FLNC
ZFPM2
LTBP2
TP53BP1
INTS1
PINK1-AS
MLLT4
unidentified
TRIM14
unidentified
unidentified
Hybridization round
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Spot number
1 0 0 1 0 0 0 0 0 0 1 0 0 1
2 0 0 0 0 0 0 0 1 0 1 0 1 1
3 0 0 1 0 0 0 1 0 0 0 0 1 1
4 0 1 0 0 1 1 0 0 0 0 0 0 0
5 1 0 0 0 0 0 0 1 0 0 0 1 0
6 0 1 0 0 1 1 0 0 0 0 1 0 0
7 0 0 1 0 0 0 0 1 1 0 1 0 0
8 0 0 0 0 1 0 0 1 1 1 0 0 0
9 0 0 0 0 0 0 0 0 0 1 0 0 0
10 0 0 0 1 1 0 1 0 0 1 0 0 0
11 0 0 0 1 0 0 0 0 0 0 0 0 0
12 0 0 1 0 0 0 0 0 0 0 0 0 0

Fig. 19B

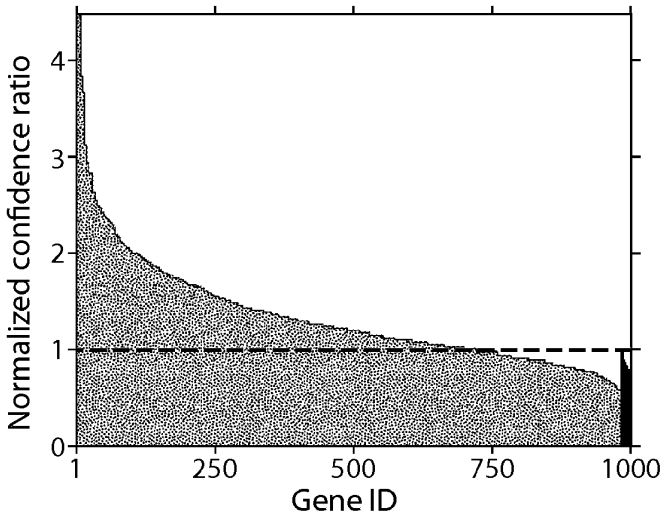


Fig. 19C

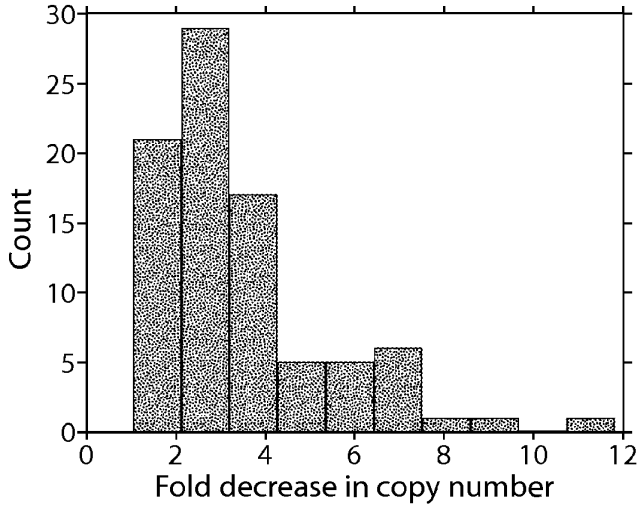


Fig. 19D

Codebook 1 for the 140-gene experiments

Genes	Codewords
THBS1	0 0 0 0 1 1 1 0 0 0 0 0 0 0 1 0
FLNA	0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 1
TNC	1 0 0 0 1 0 1 0 0 0 0 0 1 0 0 0
FBN2	0 0 1 0 1 0 1 1 0 0 0 0 0 0 0 0
KIAA1199	0 0 0 0 1 1 0 1 0 0 0 1 0 0 0 0
COL5A1	0 0 0 0 0 1 0 0 1 1 0 0 0 0 1 0
MALAT1	0 0 0 0 0 1 1 0 1 0 0 0 1 0 0 0
TLN1	0 0 0 0 1 0 0 0 0 0 1 0 0 1 1 0
FBN1	0 0 1 0 0 1 0 1 0 0 0 0 0 0 1 0
SRRM2	0 0 1 0 0 0 1 0 0 0 0 0 0 0 1 1
TEAD1	0 0 1 0 0 1 1 0 0 0 0 1 0 0 0 0
IQGAP1	0 0 0 0 0 0 1 1 0 0 0 1 0 0 1 0
MYH10	0 0 1 0 1 0 0 0 0 0 0 1 0 0 1 0
PAPPA	0 0 0 1 0 1 0 0 0 0 1 0 0 0 1 0
NR4A1	0 0 0 1 1 0 1 0 0 0 1 0 0 0 0 0
LRP1	0 0 0 0 0 1 1 0 0 0 1 0 0 1 0 0
SLC38A1	1 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0
SMARCA5	0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 1
CKAP5	0 1 0 0 0 0 1 0 0 0 1 0 0 0 1 0
PIEZO1	1 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0
VCAN	0 0 0 0 1 0 1 0 1 1 0 0 0 0 0 0
PRRC2B	0 1 0 1 0 1 1 0 0 0 0 0 0 0 0 0
SON	1 0 0 0 0 1 0 0 0 0 0 0 1 0 1 0
SLC7A11	1 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0
GPR107	0 1 0 0 0 1 0 0 0 0 0 0 0 1 1 0
KIAA1462	0 1 0 1 1 0 0 0 0 0 0 0 0 0 1 0
notarget001	0 0 0 0 1 0 0 0 1 0 0 0 1 0 1 0
blank001	0 1 0 0 1 1 0 0 0 0 1 0 0 0 0 0
ROCK2	0 1 0 0 1 0 1 0 0 0 0 0 0 1 0 0
SPTBN1	0 0 0 1 1 1 0 0 0 0 0 0 0 1 0 0
FLNC	0 0 0 1 0 0 1 0 0 0 0 0 0 0 1 1 0
PRPF8	0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 1 1
IGF2R	0 0 0 0 1 1 0 0 0 1 0 0 1 0 0 0
FOSB	1 0 0 0 0 1 1 0 0 1 0 0 0 0 0 0
UBR5	0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 1
SVEP1	0 0 0 0 0 0 1 0 0 1 0 0 1 0 1 0
CBX5	0 0 1 0 1 1 0 0 0 0 0 0 0 0 0 1
PDS5A	1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 1
NUMA1	0 0 0 0 0 0 1 0 0 1 1 0 0 0 0 1
PLXNA1	0 0 0 1 0 0 0 0 0 0 0 1 1 0 1 0

Fig. 20-1

Codebook 1 for the 140-gene experiments

Genes	Codewords
ALPK2	0 1 0 0 0 1 0 0 0 1 0 1 0 0 0 0
AFAP1	0 0 0 0 0 0 1 0 1 0 1 1 0 0 0 0
SPTAN1	0 0 1 1 1 0 0 0 0 0 0 0 1 0 0 0
DOCK7	0 1 1 0 0 0 0 0 0 0 0 0 1 0 1 0
NOTCH2	0 0 0 0 0 0 0 0 1 0 0 0 0 1 1 1
DYNC1H1	1 0 0 0 1 0 0 0 0 0 0 1 0 1 0 0
ITGA2	1 0 1 0 0 0 0 0 0 0 0 0 1 1 0
PRKDC	0 0 0 1 1 0 0 0 0 1 0 0 0 0 0 1
GTF3C1	0 0 1 0 0 1 0 0 0 0 0 0 1 1 0 0
THSD4	0 0 0 0 0 1 0 0 0 1 0 0 0 1 0 1
PTPN14	0 1 0 0 1 0 0 1 0 1 0 0 0 0 0 0
LUZP1	0 0 0 0 1 0 0 0 1 0 1 0 0 0 0 1
AMOTL1	1 0 0 0 0 0 1 1 0 0 1 0 0 0 0 0
MAN1A2	1 0 1 0 1 0 0 0 0 0 1 0 0 0 0 0
notarget002	1 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0
blank002	1 1 0 0 0 0 0 1 0 0 0 0 0 0 1 0
USP9X	0 0 0 0 1 0 0 0 0 1 1 1 0 0 0 0
CREBBP	0 1 0 0 0 0 1 0 0 0 0 0 1 0 0 1
KPNA4	0 1 1 0 0 0 1 0 0 1 0 0 0 0 0 0
EGFR	0 0 0 1 0 0 1 0 0 1 0 1 0 0 0 0
ZC3H13	0 0 0 1 0 0 1 1 0 0 0 0 1 0 0 0
TNRC6A	0 0 0 1 0 1 0 1 1 0 0 0 0 0 0 0
TPR	0 0 0 0 0 1 0 0 1 0 0 1 0 1 0 0
COL7A1	0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 1
ASCC3	0 1 0 0 0 0 1 1 1 0 0 0 0 0 0 0
USP34	0 0 0 0 0 0 0 0 0 0 1 0 1 0 1 1
CENPF	0 0 1 0 0 0 1 0 1 0 0 0 0 1 0 0
KIDINS220	0 0 0 0 0 1 0 0 0 0 1 1 1 0 0 0
SCUBE3	1 0 0 0 0 1 0 1 0 0 0 0 0 1 0 0
NRIP1	0 0 0 0 0 0 0 0 0 1 0 1 0 1 1 0
NUP98	1 0 0 1 0 0 0 0 0 0 0 0 0 0 1 1
ZFC3H1	0 0 0 0 0 0 0 1 1 0 1 0 0 0 1 0
FAM208B	0 1 0 0 0 0 0 0 0 1 0 0 0 0 1 1
RAB3B	0 0 0 1 0 0 0 1 0 1 0 0 0 0 1 0
AFF4	1 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0
PRKCA	0 0 1 0 0 0 1 0 0 0 1 0 1 0 0 0
FLNB	1 0 0 0 0 0 1 0 0 0 0 0 0 1 0 1
CHD8	1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1
ANKRD52	0 1 0 0 1 0 0 0 0 0 0 1 1 0 0 0
MKI67	0 0 0 0 0 0 1 0 0 0 0 1 1 1 0 0

Fig. 20-2

Codebook 1 for the 140-gene experiments

Genes	Codewords
DIP2C	0 1 1 0 1 0 0 0 1 0 0 0 0 0 0 0
SSH1	0 0 0 0 1 0 0 0 0 0 0 0 1 1 0 1
notarget003	0 0 1 0 0 1 0 0 1 0 1 0 0 0 0 0
blank003	1 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0
DNAJC13	0 0 0 0 1 0 0 1 1 0 0 0 0 1 0 0
DSEL	0 1 0 0 0 0 0 0 1 0 0 1 0 0 1 0
PIK3CA	0 0 0 0 0 1 0 1 0 1 1 0 0 0 0 0
MED14	1 0 0 1 0 1 0 0 0 0 0 1 0 0 0 0
HERC2	0 1 0 0 0 1 0 0 1 0 0 0 0 0 0 1
H6PD	0 0 1 1 0 0 0 0 1 0 0 0 0 0 1 0
C14orf132	0 0 0 0 0 0 0 1 0 0 0 0 1 1 1 0
SIPA1L3	0 0 0 0 0 0 1 1 0 1 0 0 0 1 0 0
C17orf51	0 0 1 1 0 1 0 0 0 1 0 0 0 0 0 0
NF1	0 0 0 1 0 0 1 0 1 0 0 0 0 0 0 1
VPS13A	1 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0
ABCA2	1 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0
PCNX	0 1 0 0 0 1 0 1 0 0 0 0 1 0 0 0
NKTR	0 0 1 0 1 0 0 0 0 1 0 0 0 1 0 0
FYCO1	0 0 0 1 1 0 0 0 1 0 0 1 0 0 0 0
ATP11A	0 0 1 0 0 0 0 0 0 1 1 0 0 0 1 0
REV3L	0 0 0 0 1 0 0 1 0 0 1 0 1 0 0 0
CASC5	0 1 0 0 0 0 0 1 0 0 1 1 0 0 0 0
AKAP11	0 1 0 1 0 0 0 0 0 0 1 0 0 1 0 0
CEP250	0 0 1 0 0 0 0 0 0 1 0 0 1 0 0 1
USP24	0 0 0 0 0 0 0 0 1 0 1 0 1 1 0 0
AGPS	1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1
CBL	1 0 1 0 0 0 0 0 0 1 0 1 0 0 0 0
RNF169	0 0 0 1 0 0 0 0 0 0 1 1 0 0 0 1
TMOD2	0 0 0 1 0 0 0 0 0 1 0 0 1 1 0 0
FASN	0 0 0 0 0 0 0 0 1 1 0 1 0 0 0 1
notarget004	1 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0
blank004	0 0 1 0 0 0 0 0 1 0 0 1 1 0 0 0
UBR2	0 1 1 0 0 0 0 0 0 0 1 0 0 0 0 1
PHIP	1 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1
GTF3C4	0 0 0 0 0 0 0 1 0 1 0 1 1 0 0 0
QSER1	1 0 0 1 0 0 0 0 0 0 1 0 1 0 0 0
VPS13D	0 0 1 0 0 0 0 1 1 1 0 0 0 0 0 0
FRY	0 0 0 1 0 0 0 0 0 1 1 1 0 0 0 0 0
DIP2B	0 1 0 1 0 0 0 0 0 1 0 0 0 1 0 0 0
HIVEP2	1 1 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0

Fig. 20-3

Codebook 1 for the 140-gene experiments

Genes	Codewords
DOPEY1	1 0 0 0 0 0 0 0 1 1 0 0 1 0 0 0
ALMS1	0 1 1 0 0 0 0 1 0 0 0 0 0 1 0 0
HEATR5B	0 0 1 1 0 0 0 1 0 0 1 0 0 0 0 0
HIVEP1	0 0 1 0 0 0 0 1 0 0 0 1 0 0 0 1
ARL10	0 1 0 1 0 0 0 1 0 0 0 0 0 0 0 1
HELZ2	0 0 0 0 0 0 0 1 1 0 0 0 1 0 0 1
POLQ	1 0 1 0 0 0 0 0 1 0 0 0 0 0 0 1
CACNA1A	1 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0
KIAA2018	0 1 0 0 0 0 0 0 0 0 0 1 0 1 0 1
LYST	1 0 0 0 0 0 0 1 1 0 0 1 0 0 0 0
ZKSCAN2	0 0 1 1 0 0 0 0 0 0 0 0 0 1 0 1
DIEXF	0 0 0 1 0 0 0 1 0 0 0 1 0 1 0 0
FZD4	1 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0
SLC5A3	0 1 1 1 0 0 0 0 0 0 0 1 0 0 0 0
KIAA1147	1 1 0 0 0 0 0 0 1 0 1 0 0 0 0 0
LMTK2	0 1 0 0 0 0 0 0 0 1 1 0 1 0 0 0
STARD9	0 0 1 0 0 0 0 0 0 0 1 1 0 1 0 0
BRCA2	0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 1
notarget005	1 0 0 0 0 0 0 0 0 1 1 0 0 1 0 0
blank005	1 1 0 0 0 0 0 0 1 0 1 0 0 0 0 0

Fig. 20-4

Codebook 2 (shuffled) for the 140-gene experiments

Genes	Codewords
THBS1	0 0 0 0 1 0 0 0 0 0 0 0 1 1 0 1
FLNA	0 0 0 0 1 1 0 0 0 1 0 0 1 0 0 0
TNC	1 0 0 0 0 0 1 0 0 0 0 0 0 1 0 1
FBN2	0 1 0 0 0 0 0 0 0 0 0 1 0 1 0 1
KIAA1199	0 0 0 0 0 0 0 0 0 0 1 0 1 0 1 1
COL5A1	0 0 0 0 0 0 0 1 0 0 0 0 1 1 1 0
MALAT1	0 0 0 0 0 0 0 0 1 0 0 0 0 1 1 1
TLN1	0 0 0 1 0 0 0 0 0 1 0 0 1 1 0 0
FBN1	1 0 0 0 1 0 0 0 0 0 0 1 0 1 0 0
SRRM2	0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 1
TEAD1	1 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0
IQGAP1	0 0 1 0 1 1 0 0 0 0 0 0 0 0 0 1
MYH10	0 0 0 1 1 0 0 0 0 1 0 0 0 0 0 1
PAPPA	0 0 0 0 1 0 0 1 0 0 1 0 1 0 0 0
NR4A1	0 0 0 0 1 0 0 0 1 0 1 0 0 0 0 1
LRP1	0 0 0 0 0 0 0 0 1 0 1 0 1 1 0 0
SLC38A1	0 0 0 0 0 1 0 0 0 1 0 0 0 1 0 1
SMARCA5	0 0 1 0 0 0 0 0 0 1 0 0 1 0 0 1
CKAP5	0 0 1 0 0 1 0 0 0 0 0 0 1 1 0 0
PIEZO1	0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 1
VCAN	0 0 0 0 0 0 1 0 0 0 0 1 1 1 0 0
PRRC2B	1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1
SON	0 0 1 0 1 0 0 0 0 1 0 0 0 1 0 0
SLC7A11	0 1 0 0 1 0 1 0 0 0 0 0 0 1 0 0
GPR107	0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 1
KIAA1462	0 0 1 1 1 0 0 0 0 0 0 0 1 0 0 0
notarget001	1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 1
blank001	1 0 0 0 1 0 1 0 0 0 0 0 1 0 0 0
ROCK2	0 1 0 0 0 0 1 0 0 0 0 0 1 0 0 1
SPTBN1	0 0 0 0 1 0 0 0 0 0 1 0 0 1 1 0
FLNC	0 0 0 0 1 0 0 0 1 0 0 0 1 0 1 0
PRPF8	0 0 0 0 0 0 0 1 1 0 0 0 1 0 0 1
IGF2R	0 0 0 1 1 1 0 0 0 0 0 0 0 1 0 0
FOSB	0 0 1 1 0 0 0 0 0 0 0 0 0 1 0 1
UBR5	0 1 0 0 1 0 0 0 0 0 0 1 1 0 0 0
SVEP1	0 0 0 0 1 0 0 1 1 0 0 0 0 1 0 0
CBX5	0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 1
PDS5A	1 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0
NUMA1	0 1 0 0 0 1 0 1 0 0 0 0 1 0 0 0
PLXNA1	0 1 0 1 0 0 0 0 0 0 1 0 0 1 0 0

Fig. 20-5

Codebook 2 (shuffled) for the 140-gene experiments

Genes	Codewords
ALPK2	0 1 1 0 0 0 0 0 0 0 0 0 0 1 0 1 0
AFAP1	1 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0 0
SPTAN1	0 0 0 0 0 0 0 0 0 0 1 0 1 0 1 1 0
DOCK7	0 0 1 0 0 0 0 0 0 0 1 1 0 1 0 0
NOTCH2	0 0 0 0 0 1 1 0 1 0 0 0 1 0 0 0
DYNC1H1	0 1 1 0 1 0 0 0 1 0 0 0 0 0 0 0
ITGA2	0 0 0 1 1 0 0 0 1 0 0 1 0 0 0 0
PRKDC	0 0 0 0 0 0 0 0 0 1 1 0 1 0 0 0 1
GTF3C1	0 0 1 0 1 0 0 0 0 0 0 1 0 0 1 0
THSD4	1 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1
PTPN14	0 0 0 0 0 0 1 1 0 1 0 0 0 1 0 0
LUZP1	1 0 1 0 0 0 0 0 0 0 0 0 0 1 1 0
AMOTL1	0 0 0 0 1 0 1 0 1 1 0 0 0 0 0 0
MAN1A2	0 1 0 1 0 0 0 1 0 0 0 0 0 0 0 1
notarget002	1 0 0 0 0 1 0 0 0 0 0 0 1 0 1 0
blank002	1 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0
USP9X	0 0 0 0 0 0 0 0 1 0 1 0 1 1 0 0 0
CREBBP	0 1 0 0 0 1 0 0 1 0 0 0 0 0 0 1
KPNA4	1 0 0 1 0 0 0 0 0 0 0 0 0 0 1 1
EGFR	1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1
ZC3H13	1 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0
TNRC6A	0 0 0 0 1 0 0 0 0 1 1 1 0 0 0 0
TPR	1 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0
COL7A1	0 0 1 0 0 0 0 1 0 0 0 1 0 0 0 1
ASCC3	0 1 0 0 1 0 0 1 0 1 0 0 0 0 0 0
USP34	0 0 1 0 0 0 1 0 0 0 1 0 1 0 0 0
CENPF	0 0 1 0 0 0 0 0 1 0 0 1 1 0 0 0
KIDINS220	0 0 0 1 0 0 0 0 0 0 0 1 1 0 0 0 1
SCUBE3	0 0 0 0 0 0 1 0 0 1 1 0 0 0 0 1
NRIP1	0 0 0 1 1 0 1 0 0 0 1 0 0 0 0 0
NUP98	0 0 0 0 0 1 0 0 0 0 1 1 1 0 0 0
ZFC3H1	0 0 0 1 0 0 1 0 1 0 0 0 0 0 0 1
FAM208B	0 1 0 1 1 0 0 0 0 0 0 0 0 0 1 0
RAB3B	0 0 0 1 0 0 0 0 0 0 0 0 1 1 0 1 0
AFF4	0 1 0 0 0 0 0 0 0 1 0 0 0 0 1 1
PRKCA	0 1 0 0 0 1 0 0 0 0 0 0 0 1 1 0
FLNB	0 1 0 0 0 0 0 0 1 1 0 0 0 1 0 0
CHD8	0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 1 1
ANKRD52	0 0 1 0 0 0 1 0 0 0 0 0 0 0 1 1
MKI67	0 1 0 0 1 1 0 0 0 0 1 0 0 0 0 0

Fig. 20-6

Codebook 2 (shuffled) for the 140-gene experiments

Genes	Codewords
DIP2C	1 0 1 0 1 0 0 0 0 0 1 0 0 0 0 0
SSH1	0 1 0 1 0 0 0 0 1 0 0 0 1 0 0 0
notarget003	0 1 1 0 0 0 0 0 0 0 1 0 0 0 0 1
blank003	0 0 0 0 1 1 1 0 0 0 0 0 0 0 1 0
DNAJC13	0 0 0 1 0 0 1 1 0 0 0 0 1 0 0 0
DSEL	0 0 0 1 0 0 0 1 0 0 0 1 0 1 0 0
PIK3CA	0 0 0 0 0 1 1 0 0 0 1 0 0 1 0 0
MED14	0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 1
HERC2	1 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0
H6PD	0 0 1 0 0 0 1 0 1 0 0 0 0 1 0 0
C14orf132	0 0 0 0 0 1 0 0 1 0 0 1 0 1 0 0
SIPA1L3	1 0 1 0 0 0 0 0 1 0 0 0 0 0 0 1
C17orf51	0 1 0 0 0 0 0 0 0 1 1 0 1 0 0 0
NF1	1 0 0 0 0 0 0 0 1 1 0 0 1 0 0 0
VPS13A	0 0 1 0 1 0 1 1 0 0 0 0 0 0 0 0
ABCA2	0 0 0 0 0 0 1 0 0 1 0 0 1 0 1 0
PCNX	0 0 0 1 0 0 1 0 0 0 0 0 0 1 1 0
NKTR	0 1 1 0 0 0 0 1 0 0 0 0 0 1 0 0
FYCO1	0 0 0 0 1 1 0 1 0 0 0 1 0 0 0 0
ATP11A	1 0 0 1 0 0 0 0 0 0 1 0 1 0 0 0
REV3L	1 0 0 0 0 1 0 1 0 0 0 0 0 1 0 0
CASC5	0 0 0 0 0 1 0 1 0 1 1 0 0 0 0 0
AKAP11	0 0 1 0 0 1 0 0 1 0 1 0 0 0 0 0
CEP250	1 0 1 0 0 0 0 0 0 1 0 1 0 0 0 0
USP24	1 1 0 1 0 0 0 0 0 1 0 0 0 0 0 0
AGPS	1 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0
CBL	0 1 0 0 0 0 1 1 1 0 0 0 0 0 0 0
RNF169	0 0 1 1 0 0 0 1 0 0 1 0 0 0 0 0
TMOD2	0 0 1 0 0 1 0 1 0 0 0 0 0 0 1 0
FASN	0 0 0 0 0 0 1 1 0 0 0 1 0 0 1 0
notarget004	0 1 1 0 0 0 1 0 0 1 0 0 0 0 0 0
blank004	1 1 0 0 0 0 0 1 0 0 0 0 0 0 1 0
UBR2	0 1 0 0 0 0 0 0 1 0 0 1 0 0 1 0
PHIP	0 0 1 0 0 1 1 0 0 0 0 1 0 0 0 0
GTF3C4	0 1 0 0 0 0 1 0 0 0 1 0 0 0 1 0
QSER1	1 0 0 0 0 0 1 1 0 0 1 0 0 0 0 0
VPS13D	1 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0
FRY	1 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0
DIP2B	0 1 0 0 0 1 0 0 0 1 0 1 0 0 0 0
HIVEP2	1 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0

Fig. 20-7

Codebook 2 (shuffled) for the 140-gene experiments

Genes	Codewords
DOPEY1	0 0 1 1 0 0 0 0 1 0 0 0 0 0 1 0
ALMS1	0 0 1 1 0 1 0 0 0 1 0 0 0 0 0 0
HEATR5B	0 0 1 0 0 0 0 1 1 1 0 0 0 0 0 0
HIVEP1	0 0 0 1 0 1 0 0 0 0 1 0 0 0 1 0
ARL10	0 0 0 0 0 1 0 0 1 1 0 0 0 0 1 0
HELZ2	0 0 0 1 0 0 0 1 0 1 0 0 0 0 1 0
POLQ	0 0 0 0 0 0 1 0 1 0 1 1 0 0 0 0
CACNA1A	1 0 0 0 0 1 1 0 0 1 0 0 0 0 0 0
KIAA2018	0 0 0 0 0 0 0 1 1 0 1 0 0 0 1 0
LYST	0 0 0 1 0 1 0 1 1 0 0 0 0 0 0 0
ZKSCAN2	0 1 0 1 0 1 1 0 0 0 0 0 0 0 0 0
DIEXF	0 1 0 0 0 0 0 1 0 0 1 1 0 0 0 0
FZD4	0 0 0 1 0 0 1 0 0 1 0 1 0 0 0 0
SLC5A3	1 0 0 0 0 0 0 1 1 0 0 1 0 0 0 0
KIAA1147	0 0 1 0 0 0 0 0 0 1 1 0 0 0 1 0
LMTK2	0 1 1 1 0 0 0 0 0 0 0 1 0 0 0 0
STARD9	0 0 0 1 0 0 0 0 1 1 1 0 0 0 0 0
BRCA2	1 0 0 1 0 1 0 0 0 0 0 1 0 0 0 0
notarget005	1 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0
blank005	1 1 0 0 0 0 0 0 1 0 1 0 0 0 0 0

Fig. 20-8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 15/42556

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C40B 30/04, 40/06; C07H 21/02; C12Q 1/68 (2015.01)
CPC - B01J 2219/00585, 2219/00722; C12Q 1/6876
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8) - C40B 30/04, 40/06; C07H 21/02; C12Q 1/68 (2015.01)
CPC - B01J 2219/00585, 2219/00722; C12Q 1/6876

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
CPC: B01J 2219/00585, 2219/00722; C12Q 1/6876 (text search)
USPC: 506/9, 16; 536/24.33; 435/6.11 (text search)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Electronic data bases: PatBase; Google Patents; Google Scholar
Search terms: nucleic acid probe, hybridization, barcode, index, next generation sequencing, read sequence, target sequence, error correction code, hamming code, fluorescence microscopy, transcriptsome

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2014/0031243 A1 (CAI et al.) 30 January 2014 (30.01.2014). Especially para [0007], [0018], [0052], [0069], [0141], [0178]	1-3, 88-90,166-171
Y	US 2010/0323348 A1 (HAMADY et al.) 23 January 2010 (23.01.2010).Especially para [0016], [0098].	1-3, 90, 166-171
Y	US 2010/0291557 A1 (LIVAK et al.) 18 November 2010 (18.11.2010). Especially para [0062], [0063]	88-90, 166-171
Y	US 2004/0053300 A1 (SODERLUND et al.) 18 March 2004 (18.03.2004). Especially para [0033], [0035], [0049], [0053]	166-171
X,P	CHEN et al. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science ePub 9 April 2015 Vol 348 No 6233 Pages 1-14. entire article.	1-3, 88-90,166-171

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 November 2015 (16.11.2015)	18 DEC 2015
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 15/42556

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 4-87,99-121,125-135,142,143,148-165,172-176
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
----Go to Extra Sheet for Continuation-----

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Claims 1-3, 88-90, 166-171

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (2)) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 15/42556

----Continuation of Box III (Lack of Unity of Invention)------

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I: Claims 1-3, 88-90, 166-171, drawn to a method of exposing a plurality of (primary) nucleic acid probes to a sample.

Group II: Claims 91-94, 144-147, drawn to a method of exposing a sample containing a plurality of (primary) nucleic acid probes to a plurality of secondary nucleic acid probes that are complementary to a segment of the primary nucleic acid probes and have a signaling entity attached.

Group III: Claims 95-98, drawn to a composition of a plurality of primary nucleic acid probes.

Group IV: Claims 122-124, 136-141, drawn to a method of forming a nucleic acid probe comprising a target sequence and one or more read sequences involving associating a plurality of targets with a plurality of target sequences and a plurality of codewords and associating a plurality of distinguishable read sequences with the plurality of codewords.

The inventions listed as Groups I-IV do not t relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special Technical Features:

Groups I and II have the special technical feature of exposing a sample to nucleic acid probes, not required by Groups III or IV.

Group I has the special technical feature of creating codewords based on the binding of the primary nucleic acid probes, not required by Group II-IV.

Group II has the special technical feature of creating codewords based on the binding of the secondary nucleic acid probes, not required by Group I, III or IV.

Group III has the special technical feature of being a composition of a plurality of primary nucleic acid probes, not required by Groups I, II, or IV.

Group IV has the special technical feature of a method of designing a primary nucleic acid probe, not required by Groups I, II or III.

Common Technical Features:

1. Groups I-IV share the common technical feature of a plurality of primary nucleic acid probes that comprise codewords.

2. Groups I-IV share the common technical feature of error-correction of a codeword(s).

3. Groups I-IV share the common technical feature of a primary nucleic acid probe comprising a target binding sequence and one or more artificial read sequences.

4. Groups I and II share the common technical feature of exposing sample to a plurality of nucleic acid probe.

However, said common technical features do not represent a contribution over the prior art, and is obvious over US 2014/0031243 A1 to CAI et al. (hereinafter "Cai") [published 30 January 2014], in view of US 2010/0323348 Al to HAMADY et al. (hereinafter "Hamady"), in further view of in view of US 2013/0261019 A1 to LIN et al. (hereinafter "Lin") and in further view of US 2010/0291557 Al to LIVAK et al. (hereinafter "Livak").

Concerning common technical feature #1 and #4, a plurality of primary nucleic acid probes that comprise codewords [e.g. 2 or more isoforms of a mRNA], and exposing a sample to a plurality of nucleic acid probes, Cai teaches (para [0007]; "In various embodiments, the invention teaches a method for characterizing one or more mRNA isoforms in a single cell, including: (i) creating a molecular barcode for each of said one or more mRNA isoforms in the cell, comprising: (a) providing two or more sets of two or more fluorophore-labeled oligonucleotide probes, wherein each set of probes are configured to hybridize with a specific region of one or more mRNA isoforms; and (b) hybridizing, within said cell, a quantity of said one or more mRNA isoforms with a quantity of said probes specific thereto, wherein each mRNA isoform that is hybridized with said fluorophore-labeled oligonucleotide probes emits two or more distinct signals, so as to create the molecular barcode; and (ii) resolving the molecular barcode").
-----continued on next sheet-----

----continued from previous sheet--------

Concerning common technical feature #2, error-correction of a codeword(s), error-correction of a codeword(s) based on, for example, different isoforms of a mRNA generating different or incomplete signals, would have been obvious based on the teaching of Cai, as set forth in common technical feature #1. Furthermore, Hamady teaches (para [0016]; The term "Hamming code" as used herein, refers an arithmetic process that identifies unique binary codes based upon inherent redundancy that are capable of correcting single bit errors. For example, a Hamming code can be matched with a nucleic acid barcode in order to screen for single nucleotide errors occurring during nucleic acid amplification. The identification of a single nucleotide error by using a Hamming code, thereby allows for the correction of the nucleic acid barcode"; para [0098]; Hamming codes may be represented by a subset of the possible codewords that are chosen from the center of multidimensional spheres (i.e., for example, hyperspheres) in a binary subspace. Single bit errors may fall within hyperspheres associated with a specific codeword and can thus be corrected").

Concerning common technical feature #3, a primary nucleic acid probe comprising a target binding sequence and one or more artificial read sequences, Lin teaches (para [0004]; The invention provides, inter alia, nucleic acid barcode probes that comprise a nucleic acid [....] The barcode probes may further comprise a target binding moiety"). Furthermore, concerning secondary probes to read a probe, Livak teaches (para [0062]) the invention comprises a binary probe and clamp composition in which the probe comprises a target-specific portion and a clamp-specific portion. The target- specific portion is capable of sequence-specific binding to a target polynucleotide sequence. A clamp comprises a probes pecific portion and a label. The clamp-specific portion of the probe and the probe -specific portion of the clamp form a duplex structure in the binary probe and clamp composition .... [0063] During a nucleic acid hybridization assay, the target polynucleotide sequence may be detected and quantitated in the sample by detection of the label on the binary composition. The clamp-specific portion of the probe is bound to the probe-specific portion of the clamp during the detection phase of the assay").

As the common technical features were known in the art at the time of the invention, they cannot be considered common special technical features that would otherwise unify the groups. The inventions lack unity with one another.

Therefore, Groups I-IV lack unity of invention under PCT Rule 13 because they do not share a same or corresponding special technical feature

Note concerning item 4: Claims 4-87, 99-121, 125-135, 142, 143, 148-165, 172-176 are multiple dependent claims and are not drafted according to the second and third sentences of PCT Rule 6.4(a).

Note: claim 51 cannot depend on itself. For the purpose of this invitation, claim 51 is interpreted as a dependent claim of claim 50.

Electronic Acknowledgement Receipt

EFS ID:	34365150
Application Number:	15329683
International Application Number:	
Confirmation Number:	7428
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Customer Number:	86110
Filer:	Tani Chen/Kimberly Martin
Filer Authorized By:	Tani Chen
Attorney Docket Number:	H0498.70509US02
Receipt Date:	20-NOV-2018
Filing Date:	27-JAN-2017
Time Stamp:	17:09:58
Application Type:	U.S. National Stage under 35 USC 371

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	H049870509US02-TRN-TC.pdf	26046 218d034e0b3a058601d6f637fef9cde46464c64c	no	1
Warnings:					

Information:					
2		H049870509US02-IDS-TC.pdf	43228 d201531c2b3c1e76baf32274b5a72c39a5ba55de	yes	6
	Multipart Description/PDF files in .zip description				
	Document Description		**Start**	**End**	
	Information Disclosure Statement (IDS) Form (SB08)		5	6	
	Transmittal Letter		1	4	
Warnings:					
Information:					
3	Other Reference-Patent/App/Search documents	H049870509EP00_EESR_mailed_02-06-2018.pdf	201640 4f6b8d9a11cf2a39f4faf3217943eb6d7c658dd4	no	7
Warnings:					
Information:					
4	Other Reference-Patent/App/Search documents	H049870509EP00_EP_OA_mailed_09-26-2018.PDF	147013 b20a7de0bcd1cf914c34bd76435e0556712ac762	no	5
Warnings:					
Information:					
5	Other Reference-Patent/App/Search documents	H049870510EP00_EESR_mailed_02-05-2018.pdf	232122 2c73917d577b92c9960939c01c8969d632723337	no	8
Warnings:					
Information:					
6	Other Reference-Patent/App/Search documents	H049870510EP00_EP_OA_mailed_09-24-2018.PDF	120835 45ba374937de072ac88ea23fdd9cd16f6056fae3	no	4
Warnings:					
Information:					
7	Other Reference-Patent/App/Search documents	H049870630WO00_ISR_mailed_02-05-2018.pdf	574734 bda9a5f9d340758ca0fa7a6e1c05a864bf2ee810	no	11
Warnings:					
Information:					

8	Other Reference-Patent/App/Search documents	H049870637WO00_Inv_to_Pay_Addtl_Fees_08-06-18.pdf	267420 8fa1a7c5ed5dd83378d49db153b0afa20d2010bd	no	2
Warnings:					
Information:					
9	Other Reference-Patent/App/Search documents	H049870637WO00_ISR_WO_mailed_10-04-2018.pdf	2976034 bb890075eaa20e9c894e423aed5d96984ce460cc	no	19
Warnings:					
Information:					
10	Other Reference-Patent/App/Search documents	H049870645WO00_ISR_mailed_02-05-2018.pdf	3200089 581a924bb06ea5fc05179ba0bcf1479eb64544d3	no	13
Warnings:					
Information:					
11	Foreign Reference	W01999031277.pdf	1589552 ba076f16cf4298a4a50631fd740731e8db1aae39	no	44
Warnings:					
Information:					
12	Foreign Reference	WO03003810A2.pdf	3648952 32465a7e1f52fc9c081858820f5ef67df518c207	no	75
Warnings:					
Information:					
13	Foreign Reference	WO03003810A3.pdf	137411 449dc5744fa0f22be5b8c47f938469fc777828bd	no	3
Warnings:					
Information:					
14	Foreign Reference	WO2008108843A2.pdf	6350779 1a49c51f581e3203cce6250e320a1c75aee39c0b	no	122
Warnings:					
Information:					

15	Foreign Reference	WO2013090360A2.pdf	3742469 47fa9d43bea49728c5f7076a1210ad90d1fb639f	no	66
Warnings:					
Information:					
16	Foreign Reference	WO2016018960A1.pdf	8608296 04a6d2819684df27294ee6909c7384f44b8196f7	no	163
Warnings:					
Information:					
17	Non Patent Literature	Binladen_et_al_PLoS_One_2007_v2_pe197_QC.pdf	1230999 1a35eae785dcb61da2afcbfe0cb5d3721c52cee6	no	9
Warnings:					
Information:					
18	Non Patent Literature	Chen_et_al_Nature_Methods_2016_v13_p679.PDF	6424110 c9f5ac4ea46bdfb5f5a02f2c3860a6dfcdc89dac	no	36
Warnings:					
Information:					
19	Non Patent Literature	Chen_et_al_Science_2015_v347_p543.PDF	5301311 625fb071a38212318af4e92529a538cdfaa7c877	no	24
Warnings:					
Information:					
20	Non Patent Literature	Darouich_et_al_Biotech_Histochem_2012_v2_p117.pdf	1155429 de63b0e1a851d4770566d446539f4e3a14101f77	no	5
Warnings:					
Information:					
21	Non Patent Literature	Tillberg_et_al_Nat_Biotechnol_2016_v34_p987.PDF	4425005 7aa748476ebac57d3a6f380e5cdae17576e35448	no	23
Warnings:					
Information:					
		Total Files Size (in bytes):	50403474		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/329,683-Conf. #7428
Filing Date	January 27, 2017
First Named Inventor	Xiaowei Zhuang
Art Unit	1634
Examiner Name	F. W. Lu
Attorney Docket Number	H0498.70509US02

ENCLOSURES (*Check all that apply*)

- [ ] Fee Transmittal Form
  - [ ] Fee Attached
- [ ] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [x] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) (please Identify below):

  Copies of Cited References
  Form PTO-1449

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Tani Chen/		
Printed name	Tani Chen, Sc.D.		
Date	November 20, 2018	Reg. No.	52,728

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: November 20, 2018 Electronic Signature for Kimberly E. Martin: /Kimberly E. Martin/

6771244.1

Docket No.: H0498.70509US02
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	Xiaowei Zhuang
Application No.:	15/329,683
Confirmation No.:	7428
Filed:	January 27, 2017
For:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
Examiner:	F. W. Lu
Art Unit:	1634

Certificate of Electronic Filing under 37 C.F.R. § 1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: November 20, 2018 Electronic Signature for : /Eileen M. MacKenzie/

RESPONSE TO RESTRICTION REQUIREMENT

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

In response to the Restriction Requirement set forth in the Office Action mailed September 24, 2018, Group I, claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78, is hereby elected, with traverse. The traversal is made on the grounds that a search and examination covering all claims would not present an undue burden to the Examiner.

Favorable action is respectfully requested. If, for any reason, the Examiner is of the opinion that a telephone conversation with the undersigned would expedite prosecution, the Examiner is kindly invited to contact the undersigned at 617-646-8000.

6641824.1

No fee is believed due with this response. However, if a fee is due, please charge Deposit Account No. 23/2825 under Docket No. H0498.70509US02 from which the undersigned is authorized to draw.

Dated: November 20, 2018

Respectfully submitted,

Electronic signature: /Tani Chen/
Tani Chen, Sc.D.
Registration No.: 52,728
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

Electronic Acknowledgement Receipt

EFS ID:	34368235
Application Number:	15329683
International Application Number:	
Confirmation Number:	7428
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Customer Number:	86110
Filer:	Tani Chen/Eileen MacKenzie
Filer Authorized By:	Tani Chen
Attorney Docket Number:	H0498.70509US02
Receipt Date:	20-NOV-2018
Filing Date:	27-JAN-2017
Time Stamp:	19:39:43
Application Type:	U.S. National Stage under 35 USC 371

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	H049870509US02-TRN-TC.pdf	26082 6763bfff898a9043bf6454c5499adf1a54da1db4	no	1
Warnings:					

Information:					
2	Response to Election / Restriction Filed	H049870509US02-ERRR-TC.pdf	24790 fbaae4b09ce328ee1f4ec3ce68758c16b64ac63b	no	2
Warnings:					
Information:					
			Total Files Size (in bytes):	50872	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

<u>New Applications Under 35 U.S.C. 111</u>
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.
<u>National Stage of an International Application under 35 U.S.C. 371</u>
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.
<u>New International Application Filed with the USPTO as a Receiving Office</u>
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/329,683-Conf. #7428
Filing Date	January 27, 2017
First Named Inventor	Xiaowei Zhuang
Art Unit	1634
Examiner Name	F. W. Lu
Attorney Docket Number	H0498.70509US02

ENCLOSURES (*Check all that apply*)

- [ ] Fee Transmittal Form
  - [ ] Fee Attached
- [ ] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [ ] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) (please Identify below):

  Response to Restriction Requirement

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Tani Chen/		
Printed name	Tani Chen, Sc.D.		
Date	November 20, 2018	Reg. No.	52,728

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: November 20, 2018 Electronic Signature for : /Eileen M. MacKenzie/

6641823.1

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/329,683	01/27/2017	Xiaowei Zhuang	H0498.70509US02	7428

86110 7590 04/04/2019

Wolf, Greenfield & Sacks, P.C. (Harvard)
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER	
LU, FRANK WEI MIN	

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
04/04/2019	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

H0498_eOfficeAction@WolfGreenfield.com
Patents_eOfficeAction@WolfGreenfield.com

Office Action Summary	Application No. 15/329,683	Applicant(s) Zhuang et al.	
	Examiner FRANK W LU	Art Unit 1634	AIA (FITF) Status Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☑ Responsive to communication(s) filed on 11/20/2018.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☐ This action is **FINAL.** 2b) ☑ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5) ☑ Claim(s) 1,7,10,18,26,31,34,44,50,71,75,77-78,95-96,99,103,122,126 and 129 is/are pending in the application.

5a) Of the above claim(s) 95-96,99,103,122,126 and 129 is/are withdrawn from consideration.

6) ☐ Claim(s) _____ is/are allowed.

7) ☑ Claim(s) 1,7,10,18,26,31,34,44,50,71,75 and 77-78 is/are rejected.

8) ☐ Claim(s) _____ is/are objected to.

9) ☐ Claim(s) _____ are subject to restriction and/or election requirement

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to **PPHfeedback@uspto.gov.**

Application Papers

10)☑ The specification is objected to by the Examiner.

11)☑ The drawing(s) filed on 1/27/2017 is/are: a)☑ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)

2) ☑ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date 4/12/2017 and 11/20/2018.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date _____.

4) ☐ Other: _____.

DETAILED ACTION

Election/Restrictions

1. Applicant's election with traverse of Group I, claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 in the reply filed on November 20, 2019 is acknowledged. The traversal is on the ground(s) that "[T]he traversal is made on the grounds that a search and examination covering all claims would not present an undue burden to the Examiner".

The above arguments have been fully considered and have not been found pervasive toward the withdrawal of the restriction requirement nor pervasive toward the relaxation of same such that all claims will be examined together because the restriction is not dependent on that "a search and examination covering all claims would not present an undue burden to the Examiner" as argued by applicant but is based on that "[G]roups I and III and Group II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features. For example, the special technical feature linking Groups I and III and Group II is not special because the composition in claim 95 of Group II is not a contribution over the prior art (eg., see Figures 1 and 4 of US 2012/0208724 Al, published on August 16, 2012)" and "[G]roups I and III do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features. For example, creating step in claim 1 of Group I is not required for Group III while determining sample codewords in claim 126 of Group III is not required for Group I" (see pages 3 and 4 of the office action mailed on September 24, 2018. Therefore, the requirement is still deemed proper and is made FINAL. Claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 will be examined.

Specification

2. The disclosure is objected to because of the following informalities: there are Figures 12A to 12C, 16A to 16E, and 20-1 to 20-8. However, Brief Description of the Drawings only describes Figures 12, 16A, and 16B and has no description of Figures 20-1 to 20-8.

Appropriate correction is required.

Claim Objections

3. Claim 18 is objected to because of the following informality: "the plurality of read sequences" should be "the one or more of the read sequences" in view of claim 10.

Appropriate correction is required.

Claim Rejections - 35 USC § 112

4. The following is a quotation of the first paragraph of 35 U.S.C. 112(a):

> (a) IN GENERAL.—The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor or joint inventor of carrying out the invention.

The following is a quotation of the first paragraph of pre-AIA 35 U.S.C. 112:

> The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor of carrying out his invention.

5. Scope of Enablement

Claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 are rejected under 35 U.S.C. 112(a) or 35 U.S.C. 112 (pre-AIA), first paragraph, because the specification, while being enabling for exposing a sample to a plurality of nucleic acid probes, does not reasonably provide

enablement for performing the methods recited in claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78. The specification does not enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to use the invention commensurate in scope with these claims.

In *In re Wands*, 858 F.2d 731,737, 8 USPQ2d 1400, 1404 (Fed. Cir. 1988) the court considered the issue of enablement in molecular biology. The Court summarized eight factors to be considered in a determination of "undue experimentation". These factors include: (a) the quantity of experimentation necessary; (b) the amount of direction or guidance presented; (c) the presence or absence of working examples; (d) the nature of the invention; (e) the state of the prior art; (f) the relative skill of those in the art; (g) the predictability of the art; and (h) the breadth of the claims. The Court also stated that although the level of skill in molecular biology is high, results of experiments in molecular biology are unpredictable.

To begin, there is no direction or guidance to show that the methods recited in claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 can be performed. While the relative skill in the art is very high (the Ph.D. degree with laboratory experience), there is no predictability whether the methods recited in claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 can be performed.

First, since claim 1 does not require that the sample contains either nucleic acids or proteins or peptides which can bind to the nucleic acid probes, if the sample only contains carbohydrates, it is unclear how binding of the nucleic acid probes within the sample can be determined. Furthermore, since the binding of the nucleic acid probes to the sample only can find whether the nucleic acid probes can interact with a material in the sample, it is unclear how codewords can be created only based on the binding of the nucleic acid probes to the sample. In

addition, since claim 1 does not indicate where a valid codeword comes from and how to perform an error correction, it is unclear that, if no match with a valid codeword is found, how a valid codeword can be formed by applying error correction to the codeword as recited in claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 and how an error-correcting code can be defined when the plurality of read sequences are distributed on the plurality of nucleic acid probes as recited in claim 18.

Second, since claim 26 does not require to separate the first secondary probe which is free in a solution from the first secondary probe bound to some of the read sequences of the nucleic acid probes, it is unclear how binding of the nucleic acid probes can be determined by the first signaling entity within the sample.

Third, since claim 31 does not require to separate the second secondary probe which is free in a solution from the second secondary probe bound to some of the read sequences of the nucleic acid probes and the first signaling entity is different from the second signaling entity, it is unclear how binding of the nucleic acid probes can be determined by the second signaling entity within the sample.

Fourth, since the first signaling entity can be any kind of signaling entity, it is unclear how any kind of first signaling entity can be inactivated prior to exposing the sample to the second secondary probe as recited in claim 34.

Fifth, since claim 44 does not require that the first signaling entity is a single emitter which can emit a nonoverlapping signal, it is unclear how the centroid of the first signaling entity can be determined using an algorithm for determining nonoverlapping single emitters.

Sixth, since claim 78 does not indicate which device is used to determine binding of the nucleic acid probes within the sample, it is unclear how the binding of the nucleic acid probes can be determined by imaging at least a portion of the sample.

In view of above discussions, the skilled artisan will have no way to predict the experimental results. Accordingly, it is concluded that undue experimentation is required to make the invention as it is claimed. These undue experimentation at least includes to test whether the methods recited in claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 can be performed.

6. The following is a quotation of 35 U.S.C. 112(b):

> (b) CONCLUSION.—The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the inventor or a joint inventor regards as the invention.

The following is a quotation of 35 U.S.C. 112 (pre-AIA), second paragraph:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

7. Claim 71 is rejected under 35 U.S.C. 112(b) or 35 U.S.C. 112 (pre-AIA), second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which the inventor or a joint inventor, or for pre-AIA the applicant regards as the invention.

8. Claim 71 is rejected as vague and indefinite because it is unclear that, based on what kind of device, binding of the nucleic acid probes within the sample can be determined at a resolution better than 300 nm. Please clarify.

Conclusion

9 No claim is allowed.

10. Papers related to this application may be submitted to Group 1600 by facsimile

transmission. Papers should be faxed to Group 1600 via the PTO Fax Center. The faxing of such papers must conform with the notices published in the Official Gazette, 1096 OG 30 (November 15, 1988), 1156 OG 61 (November 16, 1993), and 1157 OG 94 Fax Center number is (571)273-8300.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Frank Lu, Ph.D., whose telephone number is (571)272-0746. The examiner can normally be reached on Monday-Friday from 9 A.M. to 5 P.M.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Dave Nguyen, can be reached on (571)272-0731.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/FRANK W LU/
Primary Examiner, Art Unit 1634
March 31, 2019

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15/329,683	Zhuang et al.
	Examiner	Art Unit
	FRANK W LU	1634

CPC - Searched*		
Symbol	**Date**	**Examiner**
See attached file	03/31/2019	/FL/

CPC Combination Sets - Searched*		
Symbol	**Date**	**Examiner**
See attached file	03/31/2019	/FL/

US Classification - Searched*			
Class	**Subclass**	**Date**	**Examiner**
435	6.1, 6.11, 91.1	03/31/2019	/FL/
436	94, 501	03/31/2019	/FL/
536	23.1, 24.3	03/31/2019	/FL/

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

Search Notes		
Search Notes	**Date**	**Examiner**
STN and EAST (see attached file) and Inventor Searches	03/31/2019	/FL/

Interference Search			
US Class/CPC Symbol	**US Subclass/CPC Group**	**Date**	**Examiner**

/FRANK W LU/ Primary Examiner, Art Unit 1634	March 31, 2019

FORM PTO-1449/A and B (modified PTO/SB/08)	APPLICATION NO.: 15/329,683	ATTY. DOCKET NO.: H0498.70509US02
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	FILING DATE: January 27, 2017	CONFIRMATION NO.: 7428
	FIRST NAMED INVENTOR: Xiaowei Zhuang	
Sheet 1 of 2	GROUP ART UNIT: Not Yet Assigned	EXAMINER: Not Yet Assigned

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document Number	U.S. Patent Document Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		2004-0053300	A1	Soderlund et al.	03-18-2004
		2010-0291557	A1	Livak et al.	11-18-2010
		2010-0323348	A1	Hamady et al.	01-23-2010
		2010-0304994	A1	Wu et al.	12-02-2010
		2012-0035065	A1	Smolke et al.	02-09-2012
		2012-0129165	A1	Raj et al.	05-24-2012
		2013-0096014	A1	Anderson et al.	04-18-2013
		2013-0261019	A1	Lin et al.	10-03-2013
		2014-0031243	A1	Cai et al.	01-30-2014

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document Office/ Country	Foreign Patent Document Number	Foreign Patent Document Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		Invitation to Pay Additional Fees mailed October 6, 2015 for Application No. PCT/US2015/042556 (H0498.70509WO00).	
		International Search Report and Written Opinion mailed December 18, 2015 for Application No. PCT/US2015/042556 (H0498.70509WO00).	
		International Preliminary Report on Patentability mailed February 9, 2017 for Application No. PCT/US2015/042556 (H0498.70509WO00).	
		International Search Report and Written Opinion mailed October 9, 2015 for Application No. PCT/US2015/042559 (H098.70510WO00).	
		International Preliminary Report on Patentability mailed February 9, 2017 for Application No. PCT/US2015/042559 (H098.70510WO00).	
		CHEN et al., Spatially resolved, highly multiplexed RNA profiling in single cells. Science. 2015 Apr 24;348(6233):aaa6090. doi: 10.1126/science.aaa6090. Epub 2015 Apr 9.	
		FAKRUDDIN et al., Nucleic acid sequence based amplification (NASBA)-prospects and applications. Int. J. Life Sci. Pharm. Res. 2012 March; 2(1): L106-L121.	
		LUBECK et al., Single-cell in situ RNA profiling by sequential hybridization. Nature Methods. 2014 Apr; 11(4): 360-1.	
		LUBECK et al., Single-cell in situ RNA profiling by sequential hybridization. Nature Methods. 2014 Apr; 11(4): 360-1. Supplementary Information.	

EXAMINER:	DATE CONSIDERED:
/FRANK W LU/	03/30/2019

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

FORM PTO-1449/A and B (modified PTO/SB/08)				APPLICATION NO.: 15/329,683	ATTY. DOCKET NO.: H0498.70509US02
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				FILING DATE: January 27, 2017	CONFIRMATION NO.: 7428
				FIRST NAMED INVENTOR: Xiaowei Zhuang	
Sheet	2	of	2	GROUP ART UNIT: Not Yet Assigned	EXAMINER: Not Yet Assigned

		MOFFITT et al., High-performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing. Proc Natl Acad Sci U S A. 2016 Dec 13;113(50):14456-14461. Epub 2016 Nov 22.	
		MOFFITT et al., High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. Proc Natl Acad Sci U S A. 2016 Sep 27;113(39):11046-51. doi: 10.1073/pnas.1612826113. Epub 2016 Sep 13.	
		MOFFITT et al., RNA Imaging with Multiplexed Error-Robust Fluorescence In Situ Hybridization (MERFISH). Methods Enzymol. 2016;572:1-49. doi: 10.1016/bs.mie.2016.03.020. Epub 2016 April 27.	
		SHIELS et al., RNA-DNA hybrids containing damaged DNA are substrates for RNase H. Bioorg Med Chem Lett. 2001 Oct 8;11(19):2623-6.	

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

EXAMINER:	DATE CONSIDERED:
/FRANK W LU/	03/30/2019

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

The similar searches were performed in STN.

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	3	(DNA or RNA or "nucleic acid" or polynucleotide or oligonucleotide) near5 probe same (codeword or barcode) same valid same binding	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2019/03/31 17:18
L2	10	(DNA or RNA or "nucleic acid" or polynucleotide or oligonucleotide) near5 probe same (codeword or barcode)same binding same (correct or correction)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2019/03/31 17:19
L3	9	2 AND ((C12Q1/6806 OR C12Q1/6869 OR C12Q2563/179 OR C12Q2535/122 OR C12Q1/6816 OR C12Q2521/501 OR C12Q2531/125 OR C12Q2537/143 OR C12Q1/6827 OR C12Q1/6874 OR C12Q2525/161 OR C12Q2563/107 OR C12Q2565/101 OR C12Q2565/518 OR C12Q1/6837 OR C12Q1/6841 OR C12Q2565/514 OR C12Q1/6858 OR C12Q2521/107 OR C12Q2525/143 OR C12Q2525/179 OR C12Q2525/191 OR C12Q2525/307 OR C12Q2527/101 OR C12Q2531/113 OR C12Q2531/131 OR C12Q2565/102 OR C12Q1/68 OR C12Q1/6825 OR C12Q1/6839 OR C07H21/02 OR C07H21/04 OR C12N15/1093 OR C12N15/10 OR C12N15/1065 OR C12P19/34 OR G06N7/005 OR C40B30/04 OR C40B40/06 OR G01N21/64 OR G06F19/20 OR G06F19/22 OR G06F19/24 OR G16B25/00 OR G16B30/00 OR G16B40/00).CPC.)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2019/03/31 17:21
L4	2	1 AND ((C12Q2537/143 OR C12Q1/6816 OR C12Q2525/161 OR C12Q2563/179 OR C12Q1/6806 OR C12Q1/6837 OR C12Q2521/107 OR C12Q2525/143 OR C12Q2525/179 OR C12Q2565/102 OR C12Q2565/514 OR C12Q1/68 OR C12Q1/6841 OR C12Q1/6869 OR C07H21/02 OR C07H21/04 OR G06N7/005 OR C12N15/10 OR C12N15/1065 OR C12N15/1093 OR C40B30/04 OR G06F19/20 OR G06F19/22 OR G06F19/24 OR G16B25/00 OR G16B30/00 OR G16B40/00).CPC.)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2019/03/31 17:22

3/31/2019 5:25:52 PM
C:\Users\flu\Documents\EAST\Workspaces\~auto.05052017.131049.bak.wsp

FORM PTO-1449/A and B (modified PTO/SB/08)	APPLICATION NO.: 15/329,683	ATTY. DOCKET NO.: H0498.70509US02
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	FILING DATE: January 27, 2017	CONFIRMATION NO.: 7428
	FIRST NAMED INVENTOR: Xiaowei Zhuang	
Sheet 1 of 2	GROUP ART UNIT: 1634	EXAMINER: Frank Wei Min Lu

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document Number	U.S. Patent Document Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		6,524,798	B1	Goldbard et al.	02-25-2003
		7,368,265	B2	Brenner et al.	05-06-2008
		7,838,302		Zhuang et al.	11-23-2010
		8,564,792	B2	Zhuang et al.	10-22-2013
		2004-0137488	A1	Kenten et al.	07-15-2004
		2006-0051798	A1	Mirken et al.	03-09-2006
		2008-0269068	A1	Church et al.	10-30-2008
		2015-0087001	A1	Gradinaru et al.	03-26-2015
		2016-0116384	A1	Chen et al.	04-28-2016
		2016-0304952	A1	Boyden et al.	10-20-2016
		2017-0212986	A1	Zhuang et al.	07-27-2018

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document Office/ Country	Foreign Patent Document Number	Foreign Patent Document Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)
		WO	1999/031277	A1	Medical Science Systems, Inc.	06-24-1999	
		WO	03/003810	A3	Inst. Systems Biology	01-16-2003	
		WO	03/003810	A2	Inst. Systems Biology	01-16-2003	
		WO	2008/108843	A2	Gen Probe, Inc.	09-12-2008	
		WO	2013/090360	A2	President and Fellows of Harvard College	06-20-2013	
		WO	2016/018960	A1	President and Fellows of Harvard College	02-04-2016	

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		Extended European Search Report mailed February 6, 2018 for Application No. 15827358.1 (H0498.70509EP00).	
		European Office Action mailed September 26, 2018 for Application No. 15827358.1 (H0498.70509EP00).	
		Extended European Search Report mailed February 5, 2018 for Application No. 15828133.7 (H0498.70510EP00).	
		European Office Action mailed September 24, 2018 for Application No. 15828133.7 (H0498.70510EP00).	
		International Search Report and Written Opinion mailed February 5, 2018 for Application No. PCT/US17/60570 (H0498.70630WO00).	
		Invitation to Pay Additional Fees mailed August 6, 2018 for Application No. PCT/US18/34651 (H0498.70637WO00).	

EXAMINER:	DATE CONSIDERED:
/FRANK W LU/	03/30/2019

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

FORM PTO-1449/A and B (modified PTO/SB/08)				APPLICATION NO.: 15/329,683	ATTY. DOCKET NO.: H0498.70509US02
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				FILING DATE: January 27, 2017	CONFIRMATION NO.: 7428
				FIRST NAMED INVENTOR: Xiaowei Zhuang	
Sheet	2	of	2	GROUP ART UNIT: 1634	EXAMINER: Frank Wei Min Lu

		International Search Report and Written Opinion mailed October 4, 2018 for Application No. PCT/US18/34651 (H0498.70637WO00).	
		International Search Report and Written Opinion mailed February 5, 2018 for Application No. PCT/US17/60558 (H0498.70645WO00).	
		BINLADEN et al., The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PLoS One. 2007 Feb 14;2(2):e197.	
		CHEN et al., Expansion microscopy. Science. 2015 Jan; 347(6221):543-548	
		CHEN et al., Nanoscale imaging of RNA with expansion microscopy. Nature Methods. 2016; 13:679-684.	
		DAROUICH et al., Use of DOP-PCR for amplification and labeling of BAC DNA for FISH. Biotech Histochem. 2012 Feb;87(2):117-21. doi: 10.3109/10520295.2011.559175.	
		TILLBERG et al., Protein-retention expansion microscopy of cells and tissues labeled using standard flourescent proteins and antibodies. Nat. Biotechnol. 2016; 34(9):987-992.	

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

EXAMINER:	DATE CONSIDERED:
/FRANK W LU/	03/30/2019

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

FORM PTO-1449/A and B (modified PTO/SB/08)				APPLICATION NO.: 15/329,683	ATTY. DOCKET NO.: H0498.70509US02
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				FILING DATE: January 27, 2017	CONFIRMATION NO.: 7428
				FIRST NAMED INVENTOR: Xiaowei Zhuang	
Sheet	1	of	1	GROUP ART UNIT: 1634	EXAMINER: Frank Wei Min Lu

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document		Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		Number	Kind Code		
		7,776,613		Zhuang et al.	08-17-2010
		10,240,146	A1	Zhuang et al.	03-26-2019
		2016-0370295		Zhuang et al.	12-22-2016

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document			Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)
		Office/ Country	Number	Kind Code			
		WO	2008/091296	A2	President and Fellows of Harvard College	07-31-2008	
		WO	2016/018963	A9	President and Fellows of Harvard College	02-04-2016	

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		Chinese Office Action mailed January 16, 2019 for Application No. 201580052678.7 (H0498.70509CN00).	Y - Full
		International Preliminary report on Patentability mailed May 23, 2019 for Application No. PCT/US17/60570 (H0498.70630WO00).	
		International Preliminary report on Patentability mailed May 23, 2019 for Application No. PCT/US17/60558 (H0498.70645WO00).	
		LEVSKY et al., Single-cell gene expression profiling. Science. 2002 Aug 2; 297:836-840.	
		PLAYER et al., Single-copy gene detection using branched DNA (bDNA) in situ hybridization. The Journal of Histochemistry & Cytochemistry. 2001; 49(5): 603-611.	
		ZHEN et al., Poly-fish: A technique of repeated hybridizations that improves cytogenetic analysis of fetal cells in maternal blood. Prenatal Diagnosis. 1998 Mar 21; 18: 1181-1185.	

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

EXAMINER:	DATE CONSIDERED:

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 July 2008 (31.07.2008)

PCT

(10) International Publication Number
WO 2008/091296 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/US2007/017618

(22) International Filing Date: 7 August 2007 (07.08.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/836,167	7 August 2006 (07.08.2006)	US
60/836,170	8 August 2006 (08.08.2006)	US
11/605,842	29 November 2006 (29.11.2006)	US

(71) Applicant *(for all designated States except US)*: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **ZHUANG, Xiaowei** [CN/US]; 12 Oxford Street, Cambridge, MA 02138 (US). **BATES, Wilfred, Mark** [CA/US]; 88 Bay State Avenue, #2, Somerville, MA 02144 (US). **RUST, Michael, J.** [US/US]; 51 Bowen Avenue, #1, Medford, MA 02155 (US).

(74) Agent: **OYER, Timothy, J.**; Wolf, Greenfield & Sacks, P.C., Federal Reserve Plaza, 600 Atlantic Avenue, Boston, MA 02210-2206 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

(54) Title: SUB-DIFFRACTION LIMIT IMAGE RESOLUTION AND OTHER IMAGING TECHNIQUES

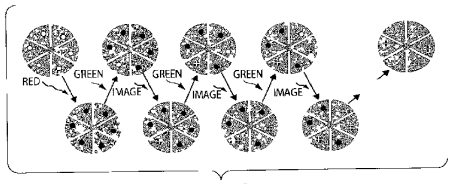


Fig. 1A

RED ROUGE
GREEN VERTE

(57) Abstract: The present invention generally relates to sub-diffraction limit image resolution and other imaging techniques. In one aspect, the invention is directed to determining and/or imaging light from two or more entities separated by a distance less than the diffraction limit of the incident light. For example, the entities may be separated by a distance of less than about 1000 run, or less than about 300 run for visible light. In one set of embodiments, the entities may be selectively activatable, i.e., one entity can be activated to produce light, without activating other entities. A first entity may be activated and determined (e.g., by determining light emitted by the entity), then a second entity may be activated and determined. The entities may be immobilized relative to each other and/or to a common entity. The emitted light may be used to determine the positions of the first and second entities, for example, using Gaussian fitting or other mathematical techniques, and in some cases, with sub-diffraction limit resolution. The methods may thus be used, for example, to determine the locations of two or more entities immobilized relative to a common entity, for example, a surface, or a biological entity such as DNA, a protein, a cell, a tissue, etc. The entities may also be determined with respect to time, for example, to determine a time-varying reaction. Other aspects of the invention relate to systems for sub-diffraction limit image resolution, computer programs and techniques for sub-diffraction limit image resolution, methods for promoting sub-diffraction limit image resolution, methods for producing photoswitchable entities, and the like.

WO 2008/091296 A2

SUB-DIFFRACTION LIMIT IMAGE RESOLUTION AND OTHER IMAGING TECHNIQUES

GOVERNMENT FUNDING

Research leading to various aspects of the present invention were sponsored, at least in part, by the Defense Advanced Research Projects Agency and the National Institutes of Health. The U.S. Government has certain rights in the invention.

RELATED APPLICATIONS

This application claims priority to all of the following according to the following recitation of priority relationships. This application claims priority to U.S. Patent Application Serial No. 11/605,842, filed November 29, 2006, entitled "Sub-Diffraction Image Resolution and other Imaging Techniques," which claims the benefit of U.S. Provisional Patent Application Serial No. 60/836,167, filed August 7, 2006, entitled "Sub-Diffraction Image Resolution"; and the benefit of U.S. Provisional Patent Application Serial No. 60/836,170, filed August 8, 2006, entitled "Sub-Diffraction Image Resolution." Each of the above is incorporated herein by reference.

FIELD OF INVENTION

The present invention generally relates to sub-diffraction limit image resolution and other imaging techniques.

BACKGROUND

Fluorescence microscopy is widely used in molecular and cell biology and other applications for non-invasive, time-resolved imaging. Despite these advantages, standard fluorescence microscopy is not useful for ultra-structural imaging, due to a resolution limit set by the diffraction of light. Several approaches have been employed to try to pass this diffraction limit, including near-field scanning optical microscopy (NSOM), multi-photon fluorescence, stimulated emission depletion (STED), reversible saturable optical linear fluorescence transition (RESOLFT), and saturated structured-illumination microscopy (SSIM), but each has certain unsatisfactory limitations. Electron microscopy is often used for high resolution imaging of biological samples, but electron microscopy uses electrons, rather than light, and is difficult to use with biological samples due to its preparation requirements. Accordingly, new techniques are

needed to harness the benefits of fluorescence microscopy for ultra-resolution imaging of biological and other samples.

SUMMARY OF THE INVENTION

The present invention generally relates to sub-diffraction limit image resolution and other imaging techniques. The subject matter of the present invention involves, in some cases, interrelated products, alternative solutions to a particular problem, and/or a plurality of different uses of one or more systems and/or articles.

The invention is a method, in one aspect. In one set of embodiments, the method includes acts of providing a first entity and a second entity separated by a distance of less than about 1000 nm, determining light emitted by the first entity, determining light emitted by the second entity, and determining the positions of the first entity and the . second entity by using the light emitted by the first entity and the light emitted by the second entity. In another set of embodiments, the method includes acts of providing a first entity and a second entity separated by a distance of less than about 1000 nm, activating the first entity but not the second entity, determining light emitted by the first entity, activating the second entity, determining light emitted by the second entity, and determining the positions of the first entity and the second entity by using the light emitted by the first entity and the light emitted by the second entity.

The method, according to yet another set of embodiments, includes acts of providing a plurality of entities able to emit light, at least some of which are separated by a distance of less than about 1000 nm, activating a fraction of the plurality of entities to emit light, determining the emitted light, deactivating the activated fraction of the plurality of entities, and repeating the acts of activating and deactivating the plurality of entities to determine the positions of the plurality of entities.

In another set of embodiments, the method includes acts of providing a first entity and a second entity separated by a distance of separation, determining light emitted by the first entity, the light emitted by the first entity having a wavelength greater than the distance of separation, determining light emitted by the second entity, and determining the positions of the first entity and the second entity by using the light emitted by the first entity and the light emitted by the second entity. The method, in yet another set of embodiments, includes acts of providing a first entity and a second entity separated by a distance of separation, activating the first entity but not the second entity,

determining light emitted by the first entity, the light emitted by the first entity having a wavelength greater than the distance of separation, activating the second entity, determining light emitted by the second entity, and determining the positions of the first entity and the second entity by using the light emitted by the first entity and the light emitted by the second entity.

In one set of embodiments, the method includes acts of providing a plurality of entities able to emit light, at least some of which are separated by a distance of separation less than the wavelength of the emitted light, activating a fraction of the plurality of entities to emit light, determining the emitted light, deactivating the activated fraction of the plurality of entities, and repeating the acts of activating and deactivating the plurality of entities to determine the positions of the plurality of entities.

The method, in another set of embodiments, includes acts of providing a first entity and a second entity separated by a distance of less than about 1000 nm where the first entity and the second entity each are immobilized relative to a common entity, determining the positions of the first entity and the second entity at a first point of time, determining the positions of the first entity and the second entity at a second point of time, and determining movement and/or structural changes of the common entity using the positions of the first and second entities at the first and second points of time. In some cases, the entities may be resolved or imaged in time. One or both of these entities may be photoactivatable or photoswitchable in some cases. The two entities may be chemically identical or distinct, for example, to allow multi-color imaging. In yet another set of embodiments, the method includes acts of providing a first entity and a second entity separated by a distance of separation where the first entity and the second entity are each immobilized relative to a common entity, determining the positions of the first entity and the second entity at a first point of time using light emitted by the first entity and light emitted by the second entity where the light emitted by the first entity has a wavelength greater than the distance of separation, determining the positions of the first entity and the second entity at a second point of time, and determining movement and/or structural changes of the common entity using the positions of the first and second entities at the first and second points of time. In some cases, the entities may be resolved or imaged in time. One or both of these entities may be photoactivatable or

photoswitchable in some cases. The two entities may be chemically identical or distinct, for example, to allow multi-color imaging.

In still another set of embodiments, the method includes acts of identifying, within a series of images in time, one or more light-emission regions, each generated by a single entity; for each light-emission region, identifying the center of the light-emission region; and for each light-emission region, reconstructing the position of the single entity generating the light-emission region at a resolution greater than the wavelength of the light emitted by the single entity. Some or all of these entities may be photoactivatable or photoswitchable. The entities may be chemically identical or distinct for example, to allow multi-color imaging.

In another aspect, the invention is directed to an article including a translation stage for a microscope having a drift of less than about 100 nm/min, and/or an article including time-modulated light sources that can be switched on and off periodically and/or in a programmed fashion, and/or an article including detectors for detecting fluorescence emission.

Still another aspect of the invention is directed to an imaging composition. The composition, according to one set of embodiments, includes a light-emitting entity, capable of being reversibly or irreversibly switched between a first state able to emit light at a first, emission wavelength and a second state that does not substantially emit light at the first wavelength. In one embodiment, the light-emitting entity comprises a first portion that is capable of emitting light at the first wavelength, and a second portion that activates the first portion upon exposure to an external stimulus, thereby causing the first portion to emit light at the first wavelength.

In another aspect, the present invention is directed to a system for performing one or more of the embodiments described herein. In another aspect, the present invention is directed to computer programs and techniques for performing one or more of the embodiments described herein. For example, one embodiment of the invention is directed to a machine-readable medium comprising a program, embodied in the medium, for causing a machine to perform a method comprising acts of identifying, within a series of images in time, one or more light-emission regions, each generated by a single entity; for each light-emission region, identifying the center of the light-emission region; and for each light-emission region, reconstructing the position of the single entity generating the

light-emission region at a resolution greater than the wavelength of the light emitted by the single entity.

Other advantages and novel features of the present invention will become apparent from the following detailed description of various non-limiting embodiments of the invention when considered in conjunction with the accompanying figures. In cases where the present specification and a document incorporated by reference include conflicting and/or inconsistent disclosure, the present specification shall control. If two or more documents incorporated by reference include conflicting and/or inconsistent disclosure with respect to each other, then the document having the later effective date shall control.

BRIEF DESCRIPTION OF THE DRAWINGS

Non-limiting embodiments of the present invention will be described by way of example with reference to the accompanying figures, which are schematic and are not intended to be drawn to scale. In the figures, each identical or nearly identical component illustrated is typically represented by a single numeral. For purposes of clarity, not every component is labeled in every figure, nor is every component of each embodiment of the invention shown where illustration is not necessary to allow those of ordinary skill in the art to understand the invention. In the figures:

Figs. Ia-Ib illustrate the principle of sub-diffraction limit resolution imaging, according to one embodiment of the invention, and a photoswitchable Cy3-Cy5 moiety;

Figs. 2a-2d illustrate the localization of molecular entities according to another embodiment of the invention;

Figs. 3a-3d illustrate the sub-diffraction limit localization of photoswitchable molecular entities according to yet another embodiment of the invention;

Fig. 4 illustrates the repeated switching of Cy3-Cy5-labeled antibody, in accordance with one embodiment of the invention;

Figs. 5a-5b illustrate drift correction in another embodiment of the invention;

Fig. 6 illustrates the repeated switching of two molecular switches, in yet another embodiment of the invention;

Figs. 7a-7e respectively illustrate the structures of Cy3, Cy5, Cy5.5, Cy7, and an example of a linked Cy3-Cy5 moiety;

Figs. 8a-8b illustrate various photoswitchable entities, according to certain embodiments of the invention;

Figs. 9a-9c illustrate absorption and emission spectra of various moieties, according to certain embodiments of the invention;

Figs. 10a-10f illustrate sub-diffraction limit imaging of cells with photoswitchable entities, according to certain embodiments of the invention;

Figs. 11a-11c illustrate multi-color, sub-diffraction limit imaging of cells with photoswitchable entities, according to certain embodiments of the invention; and

Figs. 12a-12d illustrate multi-color, sub-diffraction limit localization of multiple types of photoswitchable molecular, entities according to yet another embodiment of the invention.

DETAILED DESCRIPTION

The present invention generally relates to sub-diffraction limit image resolution and other imaging techniques. In one aspect, the invention is directed to determining and/or imaging light from two or more entities separated by a distance less than the diffraction limit of the incident light. For example, the entities may be separated by a distance of less than about 1000 nm, or less than about 300 nm for visible light. In one set of embodiments, the entities may be selectively activatable, i.e., one entity can be activated to produce light, without activating other entities. A first entity may be activated and determined (e.g., by determining light emitted by the entity), then a second entity may be activated and determined. The entities may be immobilized relative to each other and/or to a common entity. The emitted light may be used to determine the positions of the first and second entities, for example, using Gaussian fitting or other mathematical techniques, and in some cases, with sub-diffraction limit resolution. The methods may thus be used, for example, to determine the locations of two or more entities immobilized relative to (directly or indirectly, e.g., via a linker) a common entity, for example, a surface, or a biological entity such as DNA, a protein, a cell, a tissue, a biomolecular complex, etc. The entities may also be determined with respect to time, for example, to determine a time-varying reaction. Other aspects of the invention relate to systems for sub-diffraction limit image resolution, computer programs and techniques for sub-diffraction limit image resolution, methods for promoting sub-diffraction limit image resolution, methods for producing photoswitchable entities, and the like.

In various aspects of the invention, any entity able to emit light may be used. The entity may be a single molecule in some cases. Non-limiting examples of emissive entities include fluorescent entities (fluorophores) or phosphorescent entities, for example, cyanine dyes (e.g., Cy2, Cy3, Cy5, Cy5,5, Cy7, etc.) metal nanoparticles, semiconductor nanoparticles or "quantum dots," or fluorescent proteins such as GFP (Green Fluorescent Protein). Other light-emissive entities are readily known to those of ordinary skill in the art. As used herein, the term "light" generally refers to electromagnetic radiation, having any suitable wavelength (or equivalently, frequency). For instance, in some embodiments, the light may include wavelengths in the optical or visual range (for example, having a wavelength of between about 400 nm and about 700 nm, i.e., "visible light"), infrared wavelengths (for example, having a wavelength of between about 300 micrometers and 700 nm), ultraviolet wavelengths (for example, having a wavelength of between about 400 nm and about 10 nm), or the like. In certain cases, as discussed in detail below, more than one entity may be used, i.e., entities that are chemically different or distinct, for example, structurally. However, in other cases, the entities may be chemically identical or at least substantially chemically identical.

In one set of embodiments, the entity is "switchable," i.e., the entity can be switched between two or more states, at least one of which emits light having a desired wavelength. In the other state(s), the entity may emit no light, or emit light at a different wavelength. For instance, an entity may be "activated" to a first state able to produce light having a desired wavelength, and "deactivated" to a second state. An entity is "photoactivatable" if it can be activated by incident light of a suitable wavelength. As a non-limiting example, Cy5, can be switched between a fluorescent and a dark state in a controlled and reversible manner by light of different wavelengths, i.e., 633 nm red light can switch or deactivate Cy5 to a stable dark state, while 532 nm green light can switch or activate the Cy5 back to the fluorescent state. In some cases, the entity can be reversibly switched between the two or more states, e.g., upon exposure to the proper stimuli. For example, a first stimuli (e.g., a first wavelength of light) may be used to activate the switchable entity, while a second stimuli (e.g., a second wavelength of light) may be used to deactivate the switchable entity, for instance, to a non-emitting state. Any suitable method may be used to activate the entity. For example, in one embodiment, incident light of a suitable wavelength may be used to activate the entity to

emit light, i.e., the entity is "photoswitchable." Thus, the photoswitchable entity can be switched between different light-emitting or non-emitting states by incident light, e.g., of different wavelengths. The light may be monochromatic (e.g., produced using a laser) or polychromatic. In another embodiment, the entity may be activated upon stimulation by electric field and/or magnetic field. In other embodiments, the entity may be activated upon exposure to a suitable chemical environment, e.g., by adjusting the pH, or inducing a reversible chemical reaction involving the entity, etc. Similarly, any suitable method may be used to deactivate the entity, and the methods of activating and deactivating the entity need not be the same. For instance, the entity may be deactivated upon exposure to incident light of a suitable wavelength, or the entity may be deactivated by waiting a sufficient time.

Typically, a "switchable" entity can be identified by one of ordinary skill in the art by determining conditions under which an entity in a first state can emit light when exposed to an excitation wavelength, switching the entity from the first state to the second state, e.g., upon exposure to light of a switching wavelength, then showing that the entity, while in the second state can no longer emit light (or emits light at a much reduced intensity) when exposed to the excitation wavelength.

In one set of embodiments, as discussed, a switchable entity may be switched upon exposure to light. In some cases, the light used to activate the switchable entity may come from an external source (e.g., a light source such as a fluorescent light source, another light-emitting entity proximate the switchable entity, etc.). The second, light emitting entity, in some cases, may be a fluorescent entity, and in certain embodiments, the second, light-emitting entity may itself also be a switchable entity.

In some embodiments, the switchable entity includes a first, light-emitting portion (e.g,. a fluorophore), and a second portion that activates or "switches" the first portion. For example, upon exposure to light, the second portion of the switchable entity may activate the first portion, causing the first portion to emit light. Examples of activator portions include, but are not limited to, Alexa Fluor 405 (Invitrogen), Alexa 488 (Invitrogen), Cy2 (GE Healthcare), Cy3 (GE Healthcare), Cy3.5 (GE Healthcare), or Cy5 (GE Healthcare), or other suitable dyes. Examples of light-emitting portions include, but are not limited to, Cy5, Cy5.5 (GE Healthcare), or Cy7 (GE Healthcare), Alexa Fluor 647 (Invitrogen), or other suitable dyes. These may linked together, e.g.,

covalently, for example, directly, or through a linker, e.g., forming compounds such as, but not limited to, Cy5-Alexa Fluor 405, Cy5-Alexa Fluor 488, Cy5-Cy2, Cy5-Cy3, Cy5-Cy3.5, Cy5.5-Alexa Fluor 405, Cy5.5-Alexa Fluor 488, Cy5.5-Cy2, Cy5.5-Cy3, Cy5.5-Cy3.5, Cy7-Alexa Fluor 405, Cy7-Alexa Fluor 488, Cy7-Cy2, Cy7-Cy3, Cy7-Cy3.5, or Cy7-Cy5. The structures of Cy3, Cy5, Cy5,5, and Cy 7 are shown in Fig. 7, with a non-limiting example of a linked version of Cy3-Cy5 shown in Fig. 7e; those of ordinary skill in the art will be aware of the structures of these and other compounds, many of which are available commercially. The portions may be linked via a covalent bond, or by a linker, such as those described in detail below.

In certain cases, the light-emitting portion and the activator portions, when isolated from each other, may each be fluorophores, i.e., entities that can emit light of a certain, emission wavelength when exposed to a stimulus, for example, an excitation wavelength. However, when a switchable entity is formed that comprises the first fiuorophore and the second fluorophore, the first fluorophore forms a first, light-emitting portion and the second fluorophore forms an activator portion that switches that activates or "switches" the first portion in response to a stimulus. For example, the switichable entity may comprise a first fluorophore directly bonded to the second fluorophore, or the first and second entity may be connected via a linker or a common entity. Whether a pair of light-emitting portion and activator portion produces a suitable switchable entity can be tested by methods known to those of ordinary skills in the art. For example, light of various wavelength can be used to stimulate the pair and emission light from the light-emitting portion can be measured to determined wither the pair makes a suitable switch.

As a non-limiting example, Cy3 and Cy5 may be linked together to form such an entity. Such a procedure is described in more detail in the Examples, below. In this example, Cy3 is an activator portion that is able to activate Cy5, the light-emission portion. Thus, light at or near the absorption maximum (e.g., near 532 nm light for Cy3) of the activation or second portion of the entity may cause that portion to activate the first, light-emitting portion, thereby causing the first portion to emit light (e.g., near 633 nm for Cy5). As previously described, the first, light-emitting portion can subsequently be deactivated by any suitable technique (e.g., by directing 633 nm red light to the Cy5 portion of the molecule).

Accordingly, in one embodiment of the invention, a light-emitting switchable entity is provided, comprising a first, light emitting portion and a second, activation portion. The entity has a maximum emission wavelength determined by the first, light emitting portion and a maximum activation wavelength determined by the second, activation portion. Notably, the two wavelengths are not controlled by the same molecular entity, and are effectively decoupled. In some cases, the same wavelength light can be used both for activating the emitting portion to a fluorescent state and for exciting emission from and deactivating the emitting portion. Further, multiple types of switchable entities within a sample may be independently determined. For example, two switchable entities having the same activator portions but different light-emission portions can be activated by the same wavelength light applied to the sample, but emit at different wavelengths due to different light-emission portions and can be easily distinguished, even at separation distances of less than sub-diffraction limit resolutions. This can effectively yield two colors in the image. Similarly, two switchable entities having the same light-emission portions but different activator portions can be activated by different wavelength light applied to the sample, due to the different activator portions, and the light-emission portions may emit at same wavelengths and can thus be distinguished, even at separation distances of less than sub-diffraction limit resolutions. This also can effectively yield two colors in the image. When these methods are combined, four (or more) color images can be readily produced. Using this principle, multi-color imaging can be scaled up to 6 colors, 9 colors, etc., depending on the switchable and/or activator entities. This multi-color imaging principle may also be used with the imaging methods described herein to yield sub-diffraction limit resolutions, and/or used to obtained multi-color images with other imaging methods not limited to sub-diffraction limit resolutions.

In contrast, fluorescent dyes commonly used by those of ordinary skill in the art (e.g., isolated Cy5) inherently have a maximum excitation wavelength and a maximum emission wavelength that are each determined by the nature and structure of the fluorescent dye, and cannot be independently controlled. Thus, the invention, in another set of embodiments, provides a range of switchable entities where the first, light emitting portion and the second, activation portion are each independently selected. Accordingly,

in some embodiments, a larger set of colors is provided, e.g., with respect to conventional imaging methods.

Any suitable method may be used to link the first, light-emitting portion and the second, activation portion. In some cases, a linker is chosen such that the distance between the first and second portions is sufficiently close to allow the activator portion to activate the light-emitting portion as desired, e.g., whenever the light-emitting portion has been deactivated in some fashion. Typically, the portions will be separated by distances on the order of 500 nm or less, for example, less than about 300 nm, less than about 100 nm, less than about 50 nm, less than about 20 nm, less than about 10 nm, less than about 5 nm, less than about 2 nm, less than about 1 nm, etc. Examples of linkers include, but are not limited to, carbon chains (e.g., alkanes or alkenes), polymer units, or the like.

In one set of embodiments of the invention, the light-emitting switchable entity can also include other switchable fluorescent probes that do not necessarily include the two portions, such as, but not limited to, photoactivatable or photoswitchable dye molecules, natural or engineered fluorescent proteins that are photoactivatable or photoswitchable, photoactivatable or photoswitchable inorganic particles, or the like. In some cases, the switchable entity may include a first, light-emitting portion and a second, activation portion, as discussed herein.

In one set of embodiments, the switchable entity can be immobilized, e.g., covalently, with respect to a binding partner, i.e., a molecule that can undergo binding with a particular analyte. Binding partners include specific, semi-specific, and non¬specific binding partners as known to those of ordinary skill in the art. The term "specifically binds," when referring to a binding partner (e.g., protein, nucleic acid, antibody, etc.), refers to a reaction that is determinative of the presence and/or identity of one or other member of the binding pair in a mixture of heterogeneous molecules (e.g., proteins and other biologies). Thus, for example, in the case of a receptor/ligand binding pair, the ligand would specifically and/or preferentially select its receptor from a complex mixture of molecules, or vice versa. Other examples include, but are not limited to, an enzyme would specifically bind to its substrate, a nucleic acid would specifically bind to its complement, an antibody would specifically bind to its antigen. The binding may be by one or more of a variety of mechanisms including, but not

limited to ionic interactions, and/or covalent interactions, and/or hydrophobic interactions, and/or van der Waals interactions, etc. By immobilizing a switchable entity with respect to the binding partner of a target molecule or structure (e.g., DNA or a protein within a cell), the switchable entity can be used for various determination or imaging purposes. For example, a switchable entity having an amine-reactive group may be reacted with a binding partner comprising amines, for example, antibodies, proteins or enzymes. A non-limiting example of the immobilization of a switchable entity to an antibody is discussed in the Examples, below.

In some embodiments, more than one switchable entity may be used, and the entities may be the same or different. In some cases, the light emitted by a first entity and the light emitted by a second entity have the same wavelength. The entities may be activated at different times and the light from each entity may be determined separately. This allows the location of the two entities to be determined separately and, in some cases, the two entities may be spatially resolved, as discussed in detail below, even at distances of separation that are less than the light emitted by the entities or below the diffraction limit of the emitted light (i.e., "sub-diffraction limit" resolutions). In certain instances, the light emitted by a first entity and the light emitted by a second entity have different wavelengths (for example, if the first entity and the second entity are chemically different, and/or are located in different environments). The entities may be spatially resolved even at distances of separation that are less than the light emitted by the entities or below the diffraction limit of the emitted light. In certain instances, the light emitted by a first entity and the light emitted by a second entity have substantially the same wavelengths, but the two entities may be activated by light of different wavelengths and the light from each entity may be determined separately. The entities may be spatially resolved even at distances of separation that are less than the light emitted by the entities, or below the diffraction limit of the emitted light.

In some embodiments, the first, light-emitting portion and the second, activation portion as described above may not be directly covalently bonded or linked via a linker, but are each immobilized relative to a common entity. In other embodiments, two or more of the switchable entities (some of which can include, in certain cases, a first, light-emitting portion and a second, activation portion linked together directly or through a linker) may be immobilized relative to a common entity in some aspects of the invention.

The common entity in any of these embodiments may be any nonbiological entity or biological entity, for example, a cell, a tissue, a substrate, a surface, a polymer, a biological molecule such as a nucleic acid (DNA, RNA, PNA, LNA, or the like), a lipid molecule, a protein or a polypeptide, or the like, a biomolecular complex, or a biological structure, for example, an organelle, a microtubule, a clathrin-coated pit,, etc. The common entity may accordingly be determined in some fashion, e.g., imaged. As another example, two or more entities may be immobilized relative to a DNA strand or other nucleic acid strand (e.g., using antibodies, substantially complementary oligonucleotides labeled with one or more entities, chemical reactions or other techniques known to those of ordinary skill in the art), and their locations along the strand detected. In some cases, the number of base pairs (bp) separating the entities along the nucleic acid strand may be determined.

In some cases, the entities may be independently switchable, i.e., the first entity may be activated to emit light without activating a second entity. For example, if the entities are different, the methods of activating each of the first and second entities may be different (e.g., the entities may each be activated using incident light of different wavelengths). As another non-limiting example, incident light having a sufficiently weak intensity may be applied to the first and second entities such that only a subset or fraction of the entities within the incident light are activated, i.e., on a stochastic or random basis. Specific intensities for activation can be determined by those of ordinary skill in the art using no more than routine skill. By appropriately choosing the intensity of the incident light, the first entity may be activated without activating the second entity.

The second entity may be activated to emit light, and optionally, the first entity may be deactivated prior to activating the second entity. The second entity may be activated by any suitable technique, as previously described, for instance, by application of suitable incident light.

In some cases, incident light having a sufficiently weak intensity may be applied to a plurality of entities such that only a subset or fraction of the entities within the incident light are activated, e.g., on a stochastic or random basis. The amount of activation may be any suitable fraction, e.g., about 5%, about 10%, about 15%, about 20%, about 25%, about 30%, about 35%, about 40%, about 45%, about 50%, about 55%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, or

about 95% of the entities may be activated, depending on the application. For example, by appropriately choosing the intensity of the incident light, a sparse subset of the entities may be activated such that at least some of them are optically resolvable from each other and their positions can be determined. Iterative activation cycles may allow the positions of all of the entities, or a substantial fraction of the entities, to be determined. In some cases, an image with sub-diffraction limit resolution can be constructed using this information.

The two or more entities may be resolved, even at distances of separation that are less than the wavelength of the light emitted by the entities or below the diffraction limit of the emitted light, according to another aspect of the invention. The resolution of the entities may be, for instance, on the order of 1 micrometer (1000 nm) or less, as described herein. For example, if the emitted light is visible light, the resolution may be less than about 700 nm. In some cases, two (or more) entities may be resolved even if separated by a distance of less than about 500 nm, less than about 300 nm, less than about 200 nm, less than about 100 nm, less than about 80 nm, less than about 60 nm, less than about 50 nm, or less than about 40 nm. In some cases, two or more entities separated by a distance of at least about 20 nm or less than 10 nm can be resolved using embodiments of the present invention.

Light emitted by each of the switchable entities may be determined, e.g., as an image or matrix. For example, the first entity may be activated and the light emitted by the first entity determined, and the second entity may be activated (with or without deactivating the first entity) and light emitted by the second entity may be determined. The light emitted by each of the plurality of entities may be at the same or different wavelengths. Any suitable method may be used to determine the emitted light. For instance, a camera such as a CCD camera, a photodiode, a photomultiplier, or a spectrometer may be used; those of ordinary skill in the art will know of other suitable techniques. Additional non-limiting example of a suitable technique for determining light produced by the entities is discussed in the Examples, below. In some cases, multiple images (or other determinations) may be used, for example, to improve resolution and/or to reduce noise. For example, at least 2, at least 5, at least 10, at least 20, at least 25, at least 50, at least 75, at least 100, etc. images may be determined, depending on the application.

In some aspects, the light may be processed to determine the spatial positions of the two or more entities. In some cases, the positions of one or more entities, distributed within an image, may each be individually determined. In one set of embodiments, the emitted light may be processed, using Gaussian fitting or other suitable techniques, to localize the position of each of the emissive entities. Details of one suitable Gaussian fit technique are described in the Examples, below; those of ordinary skill in the art will be able to identify other suitable image-processing techniques with the benefit of the present disclosure.

Another example of an image-processing technique follows, in accordance with another embodiment of the invention. Starting with a series of images of a sample (e.g., a movie), each light-emission peak (e.g., through fluorescence, phosphorescence, etc.) is identified, and the times which the peak is present are determined. For example, a peak may be present with approximately the same intensity in a series of images with respect to time. Peaks may be fit, in some cases, to Gaussian and/or elliptical Gaussian functions to determine their centroid positions, intensities, widths, and/or ellipticities. Based on these parameters, peaks which are too dim, too wide, too skewed, etc. to yield satisfactory localization accuracy may be rejected in certain cases from further analysis. Peaks which are sporadic, moving, discontinuously present, etc. may also be discarded. By determining the center position of the peak, for example, using least-squares fitting to a 2-dimensional Gaussian function of the peak intensities, the location of the source of the peak (e.g., any entity or entities able to emit light, as discussed herein) can be determined. This process may be repeated as necessary for any or all of the peaks within the sample.

In another non-limiting example of a suitable imaging processing technique of the present invention, a series of images of a sample (e.g. a movie) may include a repetitive sequence of activation frames (e.g., in which the activation light is on) and imaging frames (e.g., in which the imaging light is on). For one or more of the imaging frames, fluorescent peaks can be identified and fit to Gaussian and/or elliptical Gaussian functions to determine their centroid positions, intensities, widths, and/or ellipticities. Based on these parameters, peaks that are too dim, too wide, too skewed, etc. to yield satisfactory localization accuracy may be rejected from further analysis. Peaks which are sporadic, moving, discontinuously present, etc. may also be discarded in some cases. By

determining the center position of the peak, for example, using least-squares fitting to a 2-dimensional Gaussian function of the peak intensities, the location of the source of the peak (e.g., any entity or entities able to emit light, as discussed herein) can be determined. Peaks appearing in an imaging frame immediately after an activation frame can be recognized as a controlled activation event and may be color-coded according to the activation laser color, in some embodiments. This process may also be repeated as necessary for any or all of the peaks within the sample.

Other image-processing techniques may also be used to facilitate determination of the entities, for example, drift correction or noise filters may be used. Generally, in drift correction, for example, a fixed point is identified (for instance, as a fiduciary marker, e.g., a fluorescent particle may be immobilized to a substrate), and movements of the fixed point (i.e., due to mechanical drift) are used to correct the determined positions of the switchable entities. In another example method for drift correction, the correlation function between images acquired in different imaging frames or activation frames can be calculated and used for drift correction. In some embodiments, the drift may be less than about 1000 nm/min, less than about 500 nm/min, less than about 300 nm/min, less than about 100 nm/min, less than about 50 nm/min, less than about 30 nm/min, less than about 20 nm/min, less than about 10 nm/min, or less than 5 nm/min. Such drift may be achieved, for example, in a microscope having a translation stage mounted for x-y positioning of the sample slide with respect to the microscope objective. The slide may be immobilized with respect to the translation stage using a suitable restraining mechanism, for example, spring loaded clips. In addition, a buffer layer may be mounted between the stage and the microscope slide. The buffer layer may further restrain drift of the slide with respect to the translation stage, for example, by preventing slippage of the slide in some fashion. The buffer layer, in one embodiment, is a rubber or polymeric film, for instance, a silicone rubber film. Accordingly, one embodiment of the invention is directed to a device, comprising a translation stage, a restraining mechanism (e.g., a spring loaded clip) attached to the translation stage able to immobilize a slide, and optionally, a buffer layer (e.g., a silicone rubber film) positioned such that a slide restrained by the restraining mechanism contacts the buffer layer.

Multiple locations on a sample may each be analyzed to determine the entities within those locations. For example, a sample may contain a plurality of various entities,

some of which are at distances of separation that are less than the wavelength of the light emitted by the entities or below the diffraction limit of the emitted light. Different locations within the sample may be determined (e.g., as different pixels within an image), and each of those locations independently analyzed to determine the entity or entities present within those locations. In some cases, the entities within each location may be determined to resolutions that are less than the wavelength of the light emitted by the entities or below the diffraction limit of the emitted light, as previously discussed.

As noted, in some embodiments, more than one type of switchable entity may be used in a sample, and the positions of each of the entities may be independently determined, in some cases, at sub-diffraction limit resolutions, e.g., to generate a multi¬color image with sub-diffraction limit resolution. For instance, by repeatedly activating and deactivating particular switchable entities in a sample, the positions of a plurality of switchable entities may be determined, in some cases, to resolutions that are less than the wavelength of the light emitted by the entities or below the diffraction limit of the emitted light.

As a non-limiting example, a sample may contain two types of light emitting portions (e.g., Cy5 and Cy5.5) and two types of activation portions (e.g., Cy3 and Cy2), for a total of four (2 x 2) types of switchable entities: Cy5-Cy3, Cy5-Cy2, Cy5.5-Cy3, and Cy5.5-Cy2. The Cy3-containing entities can be activated by applying light at a suitable wavelength without activating the Cy2-containing entities (since different activation wavelengths are required) while light emitted by the Cy5-containing molecules can be distinguished from light emitted by the Cy5.5-containing molecules (which emits light at a different wavelength). Thus, only the Cy5-Cy3 entities will be determined, while the other entities are either not activated, or are activated but the light emitted by those entities is not used. By using the above-described techniques, the positions of the Cy5-Cy3 entities may be determined within a sample, even to resolutions that are less than the wavelength of the light emitted by Cy5. Additionally, by repeating this procedure using suitable activation and emission wavelengths, the positions of the other entities may also be determined, e.g., to sub-diffraction limit resolutions. In another example, the sample may include three (or more) types of light emitting portions (e.g., Cy5, Cy5.5 and Cy7) and/or three (or more) types of activation portions (e.g., Alexa Fluor 405, Cy2 and Cy3). Accordingly, multi-color imaging (e.g., 4 colors, 6

colors, 8 colors, 9 colors, 10 colors, 12 colors, etc.) with sub-diffraction limit resolutions may be realized.

In some embodiments of the invention, the entities may also be resolved as a function of time. For example, two or more entities may be observed at various time points to determine a time-varying process, for example, a chemical reaction, cell behavior, binding of a protein or enzyme, etc. Thus, in one embodiment, the positions of two or more entities may be determined at a first point of time (e.g., as described herein), and at any number of subsequent points of time. As a specific example, if two or more entities are immobilized relative to a common entity, the common entity may then be determined as a function of time, for example, time-varying processes such as movement of the common entity, structural and/or configurational changes of the common entity, reactions involving the common entity, or the like. The time-resolved imaging may be facilitated in some cases since a switchable entity can be switched for multiple cycles, each cycle give one data point of the position of the entity.

Another aspect of the invention is directed to a computer-implemented method. For instance, a computer and/or an automated system may be provided that is able to automatically and/or repetitively perform any of the methods described herein. As used herein, "automated" devices refer to devices that are able to operate without human direction, i.e., an automated device can perform a function during a period of time after any human has finished taking any action to promote the function, e.g. by entering instructions into a computer. Typically, automated equipment can perform repetitive functions after this point in time. The processing steps may also be recorded onto a machine-readable medium in some cases.

Still another aspect of the invention is generally directed to a system able to perform one or more of the embodiments described herein. For example, the system may include a microscope, a device for activating and/or switching the entities to produce light having a desired wavelength (e.g., a laser or other light source), a device for determining the light emitted by the entities (e.g., a camera, which may include color-filtering devices, such as optical filters), and a computer for determining the spatial positions of the two or more entities. In some cases, mirrors (such as dichroic mirror or a polychroic mirror), prisms, lens, diffraction gratings, or the like may be positioned to direct light from the light source. In some cases, the light sources may be time-

modulated (e.g., by shutters, acoustic optical modulators, or the like). Thus, the light source may be one that is activatable and deactivatable in a programmed or a periodic fashion. In one embodiment, more than one light source may be used, e.g., which may be used to illuminate a sample with different wavelengths or colors. For instance, the light sources may emanate light at different frequencies, and/or color-filtering devices, such as optical filters or the like may be used to modify light coming from the light sources such that different wavelengths or colors illuminate a sample.

In some embodiments, a microscope may be configured so to collect light emitted by the switchable entities while minimizing light from other sources of fluorescence (e.g., "background noise"). In certain cases, imaging geometry such as, but not limited to, a total-internal-reflection geometry a spinning-disc confocal geometry, a scanning confocal geometry, an epi-fluorescence geometry, etc., may be used for sample excitation. In some embodiments, a thin layer or plane of the sample is exposed to excitation light, which may reduce excitation of fluorescence outside of the sample plane. A high numerical aperture lens may be used to gather the light emitted by the sample. The light may be processed, for example, using filters to remove excitation light, resulting in the collection of emission light from the sample. In some cases, the magnification factor at which the image is collected can be optimized, for example, when the edge length of each pixel of the image corresponds to the length of a standard deviation of a diffraction limited spot in the image.

In some cases, a computer may be used to control excitation of the switchable entities and the acquisition of images of the swichable entities. In one set of embodiments, a sample may be excited using light having various wavelengths and/or intensities, and the sequence of the wavelengths of light used to excite the sample may be correlated, using a computer, to the images acquired of the sample containing the switchable entities. For instance, the computer may apply light having various wavelengths and/or intensities to a sample to yield different average numbers of activated switchable elements in each region of interest (e.g., one activated entity per location, two activated entities per location, etc). In some cases, this information may be used to construct an image of the switchable entities, in some cases at sub-diffraction limit resolutions, as noted above.

In other aspects of the invention, the systems and methods described herein may also be combined with other imaging techniques known to those of ordinary skill in the art, such as high-resolution fluorescence *in situ* hybridization (FISH) or immunofluorescence imaging, live cell imaging, confocal imaging, epi-fluorescence imaging, total internal reflection fluorescence imaging, etc.

The following are incorporated herein by reference: U.S. Provisional Patent Application Serial No. 60/836,167, filed August 7, 2006, entitled "Sub-Diffraction Image Resolution"; and U.S. Provisional Patent Application Serial No. 60/836,170, filed August 8, 2006, entitled "Sub-Diffraction Image Resolution."

The following examples are intended to illustrate certain embodiments of the present invention, but do not exemplify the full scope of the invention.

EXAMPLES

This example shows a high-resolution optical microscopy, stochastic optical reconstruction microscopy ("STORM"), in which a fluorescence image is constructed from high-accuracy localization of individual fluorescent entities ("fluorophores") that are switched on and off using light of different colors, in accordance with one embodiment of the invention. The STORM imaging process in this example includes a series of imaging cycles (Fig. Ia). In each cycle, a fraction of the fluorophores in the field of view are switched on or activated, such that each of the active fluorophores is optically resolvable from the rest, i.e. their images are not overlapping. This allows the position of these fluorophores to be determined with high accuracy. Repeating this process for multiple cycles, each causing a stochastically different subset of fluorophores to be turned on or activated, enables the positions of many fluorophores to be determined and thus an overall image to be reconstructed. In these examples, an imaging resolution of approximately 20 nm is demonstrated, an improvement of more than 10 times over the resolution of conventional fluorescence microscopy, using a simple total-internal-reflection fluorescence microscope, low-power continuous-wave lasers, and a photoswitchable cyanine dye.

Fig. Ia shows that a STORM imaging sequence using a hypothetical hexameric object labeled with fluorophores can be switched between a fluorescent and a dark state by a green and a red laser, respectively. In this non-limiting illustration, all fluorophores were first switched "off to the dark state ("deactivated") by a strong red laser pulse. In

each imaging cycle, a green laser pulse was used to switch on ("activate") only a fraction of the fluorophores to give an optically resolvable set of active fluorophores. Next, under red illumination ("image"), these molecules emitted fluorescence until they were switched off, allowing their positions (indicated by white crosses) to be determined with relatively high accuracy. The overall image was then reconstructed from the fluorophore positions obtained from multiple imaging cycles.

Cyanine dye, Cy5, can be switched between a fluorescent and a dark state in a controlled and reversible manner by light of different wavelengths. Red laser light that produces fluorescent emission from Cy5 can also switch the dye to a stable dark state. Exposure to green laser light converts Cy5 back to the fluorescent state, but the recovery rate may depend in some cases on the close proximity of a secondary dye, Cy3. The Cy3-Cy5 dye pair can also be referred to as a switch. Under illumination conditions allowing single-molecule detection, it was found that such a switch, when attached or immobilized to nucleic acids or proteins, can be cycled on and off hundreds of times before permanent photobleaching occurs (Fig. Ib and Fig. 4). In Fig. Ib, a red laser (633 nm, 30 W/cm^2) was used to excite fluorescence (black line) from Cy5 and to switch Cy5 to the dark state. A green laser (532 nm, 1 W/cm^2) was used to return Cy5 to the fluorescent state. The alternating red and green line indicates the laser excitation pattern. In some cases, the recovery rate of Cy5 appeared to depend on the close proximity of Cy3.

Fig. 4 shows goat anti-mouse secondary antibody labeled with the cyanine switch exhibits photoswitching behavior similar to switch-labeled DNA. The antibody was labeled with Cy3 and Cy5 (see below) and bound to a quartz slide coated with unlabeled mouse anti-transferrin primary antibody. The trace shows the Cy5 fluorescence intensity detected from a single labeled antibody as it switched on and off until permanent photobleaching occurs after about 230 seconds. The sample was excited with a sequence of alternating green and red laser pulses (0.5 s green followed by 2 s red).

Other photoswitchable dye pairs can be constructed using similar principles. For instance, in addition to Cy5, Cy5.5, Cy7, and Alexa Fluor 647 can also be switched between a fluorescent and a dark state in a controlled and reversible manner by light of different wavelengths. When paired with Cy3, upon illumination with a red laser (633nm or 657nm), each of these four dyes (Cy5, Cy5.5, Cy7, and Alexa Fluor 647) was

initially fluorescent and then quickly switched into a non-fluorescent, dark state. A brief exposure to a green laser pulse (532 nm) led to reactivation of these dyes back to the fluorescent state (see Figs. 8a and 9). Furthermore, different activating portion can be used to activate the same emitting portion in some cases. For instance, using Cy5 as an example for the emitting portion, Cy3, Cy2, and Alexa Fluor 405 could be paired with Cy5. Upon illumination with a red laser (633 nm or 657 nm), Cy5 was observed to be quickly switched into the dark state. In this example ,the reactivation of Cy5 required different colored lasers corresponding to the absorption wavelength of the activator (Figs. 8b and 9). The Alexa 405-Cy5 pair was efficiently activated by a violet laser (405 nm), but appeared to be less sensitive to blue (457 nm) and green (532 nm) lasers. Similarly, the Cy2-Cy5 pair was more sensitive to the blue light than to violet or green light, whereas the Cy3-Cy5 pair was appeared more sensitive to the green laser.

Examples of activator portions include, but are not limited to, Alexa Fluor 405, Alexa 488, Cy2, Cy3, Cy3.5, or Cy5. Examples of light-emitting portions include, but are not limited to, Cy5, Cy5.5, Cy7, or Alexa Fluor 647. These may be linked together to form photoswitchable entities such as, but not limited to, Cy5-Alexa Fluor 405, Cy5-Alexa Fluor 488, Cy5-Cy2, Cy5-Cy3, Cy5-Cy3.5, Cy5.5-Alexa Fluor 405, Cy5.5-Alexa Fluor 488, Cy5.5-Cy2, Cy5.5-Cy3, Cy5.5-Cy3.5, Cy7-Alexa Fluor 405, Cy7-Alexa Fluor 488, Cy7-Cy2, Cy7-Cy3, Cy7-Cy3.5, or Cy7-Cy5.

In this example, the concept of STORM is shown using the Cy5-C53 switch, but any suitably optically switched fluorophores could also be used. The resolution of STORM may be limited in some cases by the accuracy with which individual switches can be localized during a switching cycle. As an example of determining the localization accuracy, a switch was attached or immobilized to a short double-stranded DNA, which was surface-immobilized at low density so that single switches were resolvable. These switches were periodically cycled on and off using green and red laser light, and the red laser also served to excite fluorescence from Cy5. No fluorescence from Cy3 was recorded in this process. The high localization accuracy of individual switches during each switching cycle defining the intrinsic resolution of STORM is shown in Fig. 2. The fluorescence image from a single switch gave a point-spread function shown in Fig. 2a. The point spread function (PSF) of the emission from a single switch on DNA during a single switching cycle. A Gaussian fit to this image (not shown) was used to localize the

centroid position of the PSF, indicating the position of the switch. The positions determined from multiple switching cycles showed a substantial spread (Fig. 2b), which was significantly reduced by correcting for sample drift over the course of the experiment (Fig. 2c, see below for additional details). Figs. 2b and 2c show the centroid positions of an individual switch determined in 20 successive imaging cycles before (Fig. 2b) and after (Fig. 2c) correction for sample drift. The scale bars are 20 nm. The standard deviation of the drift-corrected positions obtained from 20 imaging cycles was on average 8 nm for individual switches (Fig. 2d). Fig. 2d shows a histogram of the standard deviation of centroid positions. The standard deviation is determined as $(\sigma_x + \sigma_y)/2$ ((sigma-x + sigma-y)/2) for each switch using 20 imaging cycles, where σ_x (sigma-x) and σ_y (sigma-y) are the standard deviations of the centroid positions in the x and y dimensions. This histogram was constructed from 29 switches. This value gave a measure of the uncertainty in the localization of a single switch per imaging cycle. Correspondingly, the experimentally measured spread of switch positions follows a Gaussian distribution with a FWHM of 18 nm, as expected from the 8 nm standard deviation (Fig. 5). Therefore, under these imaging conditions, two switches separated by 20 nm could be resolved.

Fig. 5 shows the superimposed position distributions of 29 individual switches before (Fig. 5a) and after (Fig. 5b) correction for stage drift. The scale bar is 20 nm. To obtain these distributions, the image of each switch during a given imaging cycle was fit to a two-dimensional Gaussian function, yielding the centroid position of each switch during that cycle. Twenty imaging cycles were carried out, giving twenty measured positions per switch. The position data for each switch were realigned so that the average position of each switch was at the origin, and the measurements from different switches were then superimposed to give the overall position distribution shown in the insets. The main plots are histograms of these centroid positions. The uncorrected distribution is fairly broad and skewed asymmetrically. After correction for stage drift using fiducial markers (see below) the distribution become significantly narrower. The fit of the drift-corrected histogram to a Gaussian gave a full width at half-maximum of 18 ± 2 nm.

To demonstrate the capability of STORM to resolve fluorescent molecules in close proximity, standard samples of linear, double-stranded DNA labeled were constructed with multiple switches separated by a well-defined number of base pairs (see methods, below, for details). The DNA strands were also labeled with multiple biotins and attached to a high-density streptavidin layer (see below), increasing the likelihood of multiple attachments between the DNA and the surface so that the DNA was immobilized in the plane. DNA strands were first examined containing two switches separated by 135 base-pairs (Fig. 3a and Fig. 6), corresponding to a length of 46 nm along the DNA contour. A STORM image obtained using the imaging procedure described above (see below for details) showed two clusters of measured switch positions, indicating that the two switches were well-resolved (Fig. 3a). This image shows two separated clusters of measured switch positions (crosses), each corresponding to a single switch. The center-of-mass position of each cluster is marked by a dot. The inter-switch distances are 46 nm, 44 nm and 34 nm for these three examples. The scale bars are 20 nm. The distance between the centers of the two clouds had a mean value of 4 1 nm (Fig. 3b), in quantitative agreement with the theoretical mean (40 nm) determined using the known contour and persistence lengths of the DNA sample (see below). Fig. 3b shows a comparison between the inter-switch distances measured using STORM (grey column) and the predicted distance distribution considering the flexibility of DNA (dashed line). Fig. 3c shows STORM images of four switches attached to a double-stranded DNA, pair-wise separated by a contour length of 46 nm. The measured switch positions were clustered by an automated algorithm (see below for details) and different clusters are indicated by different symbols. The scale bars are 20 nm.

Fig. 6 is an excerpt of a fluorescence trace (black line) showing the cycling of Cy3-Cy5 switches separated by a contour length of 46 nm on dsDNA (see also Fig. 3a). Each switching cycle lasted for 5 seconds and included a brief green pulse (underlying tick marks) to activate one or two switches followed by a long red exposure (underlying bar) to excite Cy5 fluorescence and return the switches to the off state. In cycles when one switch is activated (e.g. 225-230 seconds), a single intensity level was apparent before the return to the dark state. In cycles where two switches are activated (e.g. 245-250 seconds), a "staircase" with two decreasing levels corresponding to the sequential switching off of each dye is observed. Regions where only a single switch is on and all

of the peak quality criteria are satisfied (see below) were used for single-switch localization.

Longer DNA samples labeled with four switches evenly separated by 46 nm along the contour were also imaged. The STORM images revealed four clusters of switch positions following a bent contour consistent with the persistence length of DNA and the engineered separation between the switches. These results indicated that STORM can image biological samples with sub-diffraction limit resolution and that features separated by 40 nm are well within the resolving power.

One advantage of STORM is its ability to localize a large number of switches within a diffraction-limited spot by cycling the switches on and off in a controlled manner, allowing this to be used as a general biological imaging technique. To demonstrate this capability, circular DNA plasmids were prepared and coated with RecA protein and imaged using indirect immunofluorescence with switch-labeled secondary antibody (Fig. 3d, see methods, below, for details). Here the photoswitch comprise of Cy3 and Cy5. Cy3 serve as the activating portion and Cy5 serve as the light-emitting portion. STORM images of the RecA filaments revealed their circular structure with greatly increased resolution when compared with conventional wide-field images. Parts of the filament appear to be unlabeled or kinked, which may correspond to regions of the plasmid that remain uncoated by RecA. In Fig. 3d, the top panels show indirect immunofluorescence images with switch-labeled secondary antibody taken by a total internal reflection microscope. The bottom panels are the reconstructed STORM images of the same filaments. The scale bars are 300 nm.

The presence of multiple types of photoswitchable pairs of light emitting portion and activating portion may also allow multi-color STORM imaging, e.g., as described herein. This example uses a selective activation scheme as an initial demonstration of multicolor STORM imaging. Three different DNA constructs labeled with Alexa 405-Cy5, Cy2-Cy5, or Cy3-Cy5 were mixed and immobilized on a microscope slide at a high surface density such that individual DNA molecules could not be resolved in a conventional fluorescence image. To generate a STORM image, the sample was first exposed to a red laser (633 nm) to switch off nearly all Cy5 dyes in the field of view. The sample was then periodically excited with a sequence of violet (405 nm), blue (457 nm), and green (532 nm) laser pulses, each of which activated a sparse, optically

resolvable subset of fluorophores. In between activation pulses, the sample was imaged with the red laser. The image of each activated fluorescent spot was analyzed to determine its centroid position (referred to as a localization), and a color was assigned according to the preceding activation pulse. As the same imaging laser and detection channel were used for all three dye pairs, there was no need for correction of chromatic aberration. After thousands of activation cycles, a STORM image was constructed by plotting all of the colored localizations (Figs. 12a-12c). The STORM image showed separated clusters of localizations. Each cluster corresponded to an individual DNA molecule and resulted from the repetitive localization of a single Cy5 molecule over multiple switching cycles. The majority of the localizations within each cluster displayed the same color, identifying the type of activator dye present on the DNA. There were two origins of crosstalk between colors that were identified: false activation by laser pulses of incorrect colors and non-specific activation by the red imaging laser, but both effects were quantitatively small. The localizations within each cluster approximately follow a Gaussian distribution with a full-width-half-maximum (FWHM) of 26 ± 1 nm, 25 ± 1 nm, and 24 ± 1 nm for the three color channels (Figs. 12d-12f), suggesting an imaging resolution of ~25 nm. This resolution is lower than the theoretical limit predicted from the number of photons detected.

STORM imaging can also be performed on cell samples. To demonstrate this capability, in this example, single-color immunofluorescence imaging of microtubules, which are filamentous cytoskeleton structures important for many cellular functions, were performed. BS-C-I cells were fixed and immunostained with primary antibodies against microtubules and then with a switch-labeled secondary antibody. Here, the photoswitch used comprised Cy3 and Alexa Fluor 647, with Cy3 serving as the activating portion and Alexa Fluor 647 serving as the light-emitting portion. The STORM image showed a drastic improvement in the resolution of the microtubule network as compared to the conventional fluorescence image (Fig 10). Figs. 10a, 10c, and 10e are conventional images of microtubules in regions of the cell, shown at different scales, and Figs. 10b, 10d, and 10f are STORM images of the same regions. In the regions where microtubules were densely packed and undefined in the conventional image, individual microtubule filaments were resolved by STORM (Figs. 10C-F). Whole-cell STORM images, including ~ 10^6 single-molecule localizations, were

acquired in 2 to 30 minutes. Super-resolution microtubule structures began to emerge after only about 10 to 20 seconds of STORM imaging, although this was not optimized. The effective imaging resolution was affected both by the intrinsic dye localization accuracy and the size of the antibody labels. Improvement in the effective resolution may be achieved by using direct immunofluorescence staining with dye-labeled primary antibodies or Fab fragments.

Using different photoswitchable pairs, multi-color STORM imaging can also be performed on cell samples. To demonstrate this capability, microtubules and clathrin-coated pits (CCPs), cellular structures used for receptor-mediated endocytosis, were simultaneously imaged. The microtubules and clathrin were immunostained with primary antibodies and then with switch-labeled secondary antibodies. Here, the photoswitches comprised Cy2 and Alexa Fluor 647 for microtubule imaging, and Cy3 and Alexa Fluor 647 for clathrin imaging, with Cy2 or Cy3 serving as the activating portion and Alexa Fluor 647 serving as the light-emitting portion. The 457 nm and 532 nm lasers were used to selectively activate the two pairs. Crosstalk between the two color channels due to false and non-specific activations were subtracted from the image after statistical analysis. Fig. 11 shows two-color super-resolution STORM images of microtubules and clathrin-coated pits. Fig. 11 shows the two-color STORM image presented at different scales. The green channel (457 nm activation) revealed filamentous structures for microtubules (long thin structures in these images). The red channel reveals predominantly spherical structures in these images, which were the clathrin-coated pits and vesicles.

In summary, this example demonstrates that STORM is capable of imaging biological structures with sub-diffraction limit resolution. The resolution of the technique is not limited to the wavelength of light. For the Cy3-Cy5 switch, approximately 3,000 photons were detected per switching cycle, independent of red laser intensity (data not shown), predicting a theoretical localization accuracy of 4 nm. The difference between this theoretical prediction and the measured accuracy of 8 nm is relatively small, and the discrepancy may be due to imperfect correction of stage drift and aberration due to focus drift in the measurements. This measured localization accuracy corresponds to an imaging resolution of approximately 20 nm (full width half

maximum, FWHM). Indeed, fluorescent switches separated by -40 nm were readily resolved (Fig. 3).

The cyanine switches could be turned on and off reliably for hundreds of cycles before photobleaching, allowing STORM to be used for resolving structures with many fluorophores in a potentially time-resolved manner. The circular structure of RecA filaments containing 10-20 switches and microtubule and clathrin-coated pits structures in cells could be resolved in a few minutes or less. The imaging speed may be improved, for instance, by increasing the switching rate through stronger excitation or fluorophores with faster switching kinetics, by using a camera with a faster frame rate, by using a fast camera reading scheme, such as binning pixels or only reading out a fraction of the pixels, or by other technique. STORM is thus a valuable tool for high-resolution imaging of biological or nonbiological samples. The STORM concept is also applicable to other photoswitchable fluorophores and fluorescent proteins, which will potentially allow high-resolution live-cell imaging with endogenous labels, and any other imaging applications in which sub-diffraction limit image resolution is desired.

For additional details and examples, see W. M. Bates, *et al.,* "Multicolor Super-resolution Imaging with Photo-switchable Fluorescent Probes" *Science* (in press), or M. J. Rust, *et al.*, "Sub-diffraction-limit imaging by stochastic reconstruction optical microscopy (STORM)," *Nature Methods* 3, 793-795 (2006).

Following are methods useful in the above description.

Preparation of switch-labeled DNA constructs. Biotinylated and/or amine-modified DNA oligonucleotides ("oligos") were purchased, PAGE purified, from Operon. Amine-modified oligos were labeled with amine reactive Cyanine dyes (Amersham Bioscience) or Alexa dyes (Invitrogen) post-synthetically following the protocol provided by the manufacturer. The dye-labeled oligos were purified using reverse phase HPLC. Complementary strands of DNA were annealed to form biotinylated double-stranded DNA (dsDNA) by mixing equimolar amounts of the two complementary strands in 10 niM Tris-Cl (pH 7.5), 50 mM NaCl, heating to 90 ^{0}C for two minutes, and then allowing the mixture to cool to room temperature in a heat block over a period of one hour.

Biotinylated dsDNA of varying lengths having an intra-switch distance of 135 bp were constructed by annealing complimentary oligos as described above to form three

different 45 bp dsDNA segments denoted A, B, and C, followed by a ligation reaction. The oligos were designed with specific sticky ends which only permit A to ligate to B, B to ligate to C and C to ligate to A, reading in the 5' to 3' direction. One oligo (A) contained amine-reactive sites three base pairs apart on opposite strands which were specifically labeled with Cy3 and Cy5 prior to annealing to form the optical switch. Oligos B and C contain two internal biotin modifications per strand to facilitate multivalent linkage to the streptavidin surface. After annealing, the oligos were mixed at equal concentrations and ligated overnight using T4 ligase (New England Biolabs). The resulting ligation product was purified on a 1.5% agarose gel to select bands containing the desired concatamerized dsDNA length.

Preparation of switch-labeled antibodies. Goat antimouse IgG secondary antibodies (Abeam or Invitrogen) and goat anti-rabbit antibody (Abeam) were labeled non-specifically with amine-reactive Cy2 or Cy3 (to serve as an activating portion) and Cy5 or Alexa 647 (to serve as a light-emitting portion). The average dye-to-antibody ratio was ~2:1 for Cy3 and ~0.1:1 for Cy5 in the case of Rec A imaging. In the case of microtubule and clathrin imaging, dye-to-antibody ratio was ~2:1 for Cy3 or Cy2 and ~0.4:1 for Alexa Fluor 647. Various Cy5-to-antibody or Alexa 647-to-antibody ratios may be used for STORM imaging, for example, ratios up to or greater than 0.8. Various Cy3-to-antibody ratios and Cy2-to-antibody ratios can also be used for efficiently STORM imaging. The imaging quality appears not to be very sensitive the Cy3 (or Cy2)-to-antibody ratio. These labeling ratios were chosen to minimize the fraction of antibodies labeled with more than one Cy5 molecule, due to the inefficient switching observed for antibodies labeled with multiple Cy5. The presence of more than one Cy3 molecule on a single antibody did not interfere with switching. With this labeling ratio, a significant fraction of the secondary antibody did not carry Cy5, and were not observed. This lower density of labels is typically not a problem for indirect immunofluorescence imaging and, in particular, did not prevent resolution of the circular structure of the RecA-plasmid filaments or microtubules.

Preparation of RecA filaments. Biotinylated RecA was prepared by reacting purified recombinant RecA (New England Biolabs) with amine-reactive biotin-XX (Invitrogen) in 0.1 M carbonate buffer at pH 8.3. The resulting biotinylated protein was purified on a NAP-5 size exclusion column (Amersham). RecA filaments were formed

on ΦXRF-II (Phi-XRF) plasmid DNA (New England Biolabs) by incubating RecA (25% biotinylated:75% unbiotinylated, concentration of 80 micrograms/mL) with plasmid DNA (2 micrograms/mL) in 10 mM Tris buffer at pH 7.0, 100 mM NaCl, 7 mM $MgCl_2$, and 0.8 mg/mL ATP-γ-S (ATP-gamma-S) for 1 hour at 37 ^{0}C. The resulting RecA-DNA filaments were stored at 4 ^{0}C and used for imaging the same day.

Microscope slide preparation. Quartz microscope slides (G. Finkenbeiner) were cleaned using Alconox detergent, followed by sonication for 15 minutes in acetone, 1 M aqueous KOH, ethanol, 1 M aqueous KOH, sequentially. Slides being prepared for lipid bilayers were submerged into a 5% HF solution for 2 hours after the second KOH step. Finally, the slides were rinsed with deionized water, and flame dried. The flow channels were prepared using two pieces of double-sided adhesive tape (3M) and covered with a No. 1.5 glass coverslip (VWR).

Oxygen scavenging system. All imaging buffers were supplemented with the oxygen scavenging system, which included 10% (w/v) glucose (Sigma), 0.1% (v/v) beta-mercaptoethanol (Sigma), 500 micrograms/mL glucose oxidase (Sigma), and 10 micrograms/mL catalase (Roche). The oxygen scavenging system was important for reliable photoswitching of the fluorophores. As low as 0.01% of beta-mercaptoethanol can be used in some instances for efficiently photoswitching of the cyanine dyes. The beta-mercaptoethanol can also be replaced by other reducing reagents such as glutathione and cysteins, in other embodiments.

Surface-immobilization of DNA and antibodies. To immobilize the labeled dsDNA on a surface, quartz microscope slides (G. Finkenbeiner) were cleaned using Alconox detergent, sonicated in 1 M KOH, ethanol, and 1 M KOH sequentially before being rinsed with MilliQ water and flame dried. A biotinylated bovine serum albumin (BSA, Sigma) solution (1.0 mg/mL) was first added to the slides, followed by 0.25 mg/mL streptavidin (Invitrogen), and finally the DNA sample at a low concentration (~ 30 pM) in order to obtain a low surface density of DNA molecules such that individual molecules were well separated and optically resolvable. The slides were rinsed prior to the addition of each reagent.

To immobilize DNA constructs for the 3-color STORM imaging, three different DNA constructs, each labeled with an Alexa 405-Cy5 pair, a Cy2-Cy5 pair, or a Cy3-Cy5 pair were mixed in solution and co-immobilized onto a quartz slide as described

above. A concentration of 500 pM of DNA was used to reach a high surface density of immobilized molecules.

To immobilize the DNA sample labeled with multiple switches on a quartz slide, a lipid bilayer was first formed on the slide by flowing in liposomes of egg PC and 5% biotin-PE (Avanti). The liposomes were formed according to the manufacturer's instructions, extruded through a 0.05 micrometer filter membrane at a concentration of 5 mg lipids/mL in DI water, and then mixed with a 1:1 ratio with a buffer containing 10 mM Tris at pH 8.0, 100 mM NaCl immediately before use. After a 2 hour incubation the bilayer was rinsed extensively with 50 mM Tris at pH 8.0, 10 mM NaCl, and the bilayer was incubated with streptavidin (0.25 mg/mL, Invitrogen) for 30 minutes. After extensive washing, the streptavidin surface was crosslinked in 4% v/v formaldehyde in PBS for 1 hour and rinsed with Tris buffer before allowing the biotinylated, switch-labeled DNA to bind. DNA was imaged in Tris buffer (50 mM Tris-Cl at pH 7.5, 10 mM NaCl) with the oxygen scavenging system described above.

To immobilized switch-labeled antibodies on a surface, the quartz slides were cleaned as described and incubated for 5 minutes with mouse anti-transferrin IgG (Abeam) allowing it to bind non-specifically. The slide was then incubated with the labeled secondary antibodies in PBS buffer containing 3% bovine serum albumin (BSA) for 10 min. The buffer was replaced with 50 mM Tris, 10 mM NaCl, pH 7.5 containing the oxygen scavenging system for imaging.

Immunofluorescence imaging of RecA-dsDNA filaments. Biotinylated RecA-dsDNA filaments were attached via streptavidin linkages to a quartz slide non-specifically coated with biotinylated BSA. The surface was then washed with 50 mM Tris at pH 7.0, 100 mM NaCl, 7 mM $MgCl_2$, 3% BSA w/v (block buffer) and incubated for 30 minutes to block the surface against non-specific antibody binding. The slide was then incubated in this block buffer containing monoclonal mouse antibody against RecA (Stressgen) at 2 microgram/mL for 1 hour. After extensive washing with block buffer, the slide was incubated in the block buffer containing switch-labeled secondary antibody at 0.3 microgram/mL for 1 hour. Finally, the sample was washed and imaged in 50 mM Tris at pH 7.5, 100 mM NaCl, 7 mM $MgCl_2$, supplemented with the oxygen scavenging system as described above.

Immunofluorescence imaging of microtubules and clathrin-coated pits in cells. Green monkey kidney BS-C-I cells were plated in LabTek II 8 well chambered coverglass (Nunc) at a density of 30k per well. After 16 to 24 hr, they were rinsed with phosphate buffered saline (PBS) buffer, fixed with 3% formaldehyde, and 0.1% glutaraldehyde at room temperature in PBS for 10 minutes, and quenched with 0.1% sodium borohydride in PBS for 7 minutes to reduce the unreacted aldehyde groups and fluorescent products formed during fixation. The sodium borohydride solution was prepared immediately before use to avoid hydrolysis. The fixed sample was permeablized in blocking buffer (3% BSA, 0.5% Triton X-100 in PBS) for 10 min, stained with one or both of the primary antibodies against tubulin and clathrin (2.5 micrograms/mL mouse anti-beta (β) tubulin, ATNOl from Cytoskeleton and/or 2 micrograms/mL rabbit anti-clathrin heavy chain, ab21679 from Abeam) for 30 min in blocking buffer. The sample was then rinsed with washing buffer (0.2% BSA, 0.1% Triton X-100 in PBS) three times. Corresponding secondary antibodies labeled with photoswitchable probes (2.5 micrograms/mL) were added to the sample in blocking buffer and then thoroughly rinsed after 30 minutes. Cell imaging was performed in a standard imaging buffer that contained 50 mM Tris, pH 7.5, 10 mM NaCl, 0.5 mg/mL glucose oxidase (Sigma, G2133), 40 micrograms/mL catalase (Roche Applied Science, 106810), 10% (w/v) glucose and 1% (v/v) beta-mercaptoethanol. Beta-mercaptoethanol was found to be important for the observed photoswitching behavior of Cy5, Cy5.5, and Cy7, but even a low concentration of beta-mercaptoethanol (as low as 0.02% v/v, or potentially lower) supported photoswitching. Beta-mercaptoethanol at low concentrations (0.1%) was compatible with live cell imaging. Photoswitching was also observed when beta-mercaptoethanol was replaced with cysteine (100 mM), which was also compatible with live cell imaging. Glucose oxidase was used as an oxygen scavenger system to increase the photostablity of the cyanine dyes, and cells were viable at the reported glucose oxidase concentration for at least 30 minutes.

Goat anti-mouse antibody (Invitrogen) and goat anti-rabbit antibody (Abeam) were each labeled with a mixture of amine-reactive activators and reporters. Alexa 405, Cy2, and Cy3 were used as the activating portion of the photoswitch. Alexa 647 (Invitrogen), which has very similar structural and optical properties as Cy5, was used as the light-emitting portion of the photoswitch. The concentrations of the reactive dyes

were controlled such that each antibody had, on average, two activating dyes and 0.3 - 0.4 light-emitting dyes. *Imaging procedures.* Photoswitch characterization, DNA concatamer imaging and RecA-dsDNA imaging were performed with an Olympus 1X71 microscope. Single-molecule imaging was conducted in the prism-type total internal-reflection fluorescence (TIRF) imaging geometry. The samples were excited with a 657 nm laser, and activated with a 532 nm, a 457 nm or a 405nmlaser. The fluorescence emission of the dyes was collected with a N.A. 1.25 6OX water immersion objective, and imaged onto an electron multiplying CCD camera (Andor Ixon DV897) after passing through a 665 nm long pass filter (Chroma). To track motion of the sample stage, 200 nm red fluorescent polystyrene beads (Invitrogen, F-8810) were added to the slide in Tris buffer containing 10 mM $MgCl_2$ and allowed to bind to the surface. Data was acquired using custom data acquisition software written in Labview, which enabled sequences of alternating red and green laser excitation pulses to be applied to the sample, switching the dyes on and off. Laser excitation was synchronized with the camera exposure to 1 ms accuracy.

Imaging experiments on DNA concatamers typically included of 60 laser pulse cycles to activate and deactivate the switches, where each cycle lasted for 5 s, so that the final image took 5 minutes to acquire. The multiple cluster configuration of the centroid position distribution is typically apparent within 2 minutes. The experiments on antibody-labeled RecA-dsDNA filaments included 70 switch cycles, each lasting 10 s. The ring shape of the RecA-dsDNA filaments typically became evident within 2 minutes, although the position of every switch present in the sample had not yet been identified by this time. As the rate of switching depended linearly on the excitation intensity, the cycling time, and hence the overall imaging time, could be shortened without substantially affecting the localization accuracy by increasing the red laser intensity. The green laser intensity was chosen so that typically 1-3 switches were switched on during the green pulse and, in most cases, all activated switched were turned off during the red phase of the cycle.

Raw images of RecA-dsDNA for comparison with STORM reconstructed images were formed by taking the maximum fluorescence value recorded for each pixel throughout the imaging sequence so that each switch would be equally represented regardless of the amount of time it spent in the fluorescent or dark states.

Three-color STORM imaging of the DNA sample was performed on an Olympus 1X71 inverted microscope in the prism-type TIRP configuration. A 633 nm HeNe laser was used as the imaging laser and the violet (405 nm), blue (457 nm), and green (532 nm) lasers were used as the activation light sources. The sample was first exposed to the red imaging light to switch off nearly all Cy5 dyes in the field of view. Then the sample was periodically activated with a sequence of violet, blue, and green laser pulses each of which switched on a sparse, optically resolvable subset of fluorophores which were then imaged with the red laser. Fluorescence from these probes was detected with a CCD camera after passing through a long pass emission filter (Chroma, HQ645LP). During the STORM data acquisition, the camera recorded the fluorescence signal at a constant frame rate of 19 Hz. In each switching cycle, one of the activation lasers was turned on for 1 frame, followed by 9 frames of illumination with the red imaging laser.

STORM imaging of cells was performed on the Olympus 1X71 microscope with an objective-type TIRF imaging configuration. A custom polychroic beamsplitter (z458/514/647rpc, Chroma) reflected the excitation laser light onto the sample through an objective (10Ox oil, NA 1.4, UPlanSApo, Olympus), and fluorescence emission from the sample was collected by the same objective. Emitted light was filtered with two stacked dual-band emission filters (51007m, Chroma, and 595-700DBEM, Omega Optical) before being imaged on the EMCCD camera. The use of a dual-band emitter enables fluorescence from Cy3 to be collected in addition to the fluorescence of the reporter dyes. Cy3 fluorescence collected during frames in which the green activation laser was on was used for drift correction purposes and to generate the conventional fluorescence image. For single-color STORM imaging with Alexa 647 as the light emitting portion of the photoswitch and Cy3 as the activating portion, the red laser (657 nm) was used for imaging and green (532 nm) laser pulses were for activation. For two-color STORM imaging with Cy2 and Cy3 as the activating portion, alternating blue and green (457 and 532 nm) laser pulses were used for activation. Images were acquired at a frame rate of 19 Hz. In each switching cycle, one of the activation lasers was turned on for 1 frame, followed by 9 frames of illumination with the red imaging laser. Typical laser powers used for STORM imaging were 40 mW for the red laser and 2 microwatts for each of the activation *lascrs.Image analysis.* In one example of image analysis, fluorescent structures in an averaged image were first isolated in 13 x 13 pixel square fitting window for data analysis. In a given window, the total fluorescence intensity was integrated in each frame to produce a fluorescence time trace (see Fig. 6). Several criteria were used to ensure high accuracy localization of single switches:

(1) In each switching cycle, only regions where a single switch is on for at least three frames (0.3 seconds) were used for localization analysis (see Fig. 6).

(2) The fluorescence images within these regions were fit by nonlinear least-squares regression to a continuous ellipsoidal Gaussian:

$$I(x,y) = A + I_0 e^{\left[-(x'/a)^2-(y'/b)^2\right]/2}$$

where:

$$x' = (x-x_0)\cos\theta - (y-y_0)\sin\theta$$
$$y' = (x-x_0)\sin\theta + (y-y_0)\cos\theta$$

Here, A is the background fluorescence level, I_o is the amplitude of the peak, a and b reflects the widths of the Gaussian distribution along the x and y directions, x_o and y_o describe the center coordinates of the peak, and #(theta) is the tilt angle of the ellipse relative to the pixel edges. Based on this fit, the peak ellipticity defined as $|2(\alpha - b)/(a + b)|$ was computed. If this ellipticity exceeded 15%, indicating poor image quality or the possible presence of multiple active switches, the region was rejected from the analysis.

(3) The total number of counts collected in the peak was calculated as $2\pi ab I_o$ ($2 \text{ pi } ab I_0$) and then converted to photoelectrons, and thus the number of photons detected, using the camera manufacturer's calibrated curve for the electron multiplication and ADC gain settings used during imaging. If the total number of photoelectrons in the peak was less than 2,000, the region was rejected due to insufficient statistics to achieve high localization accuracy.

The regions of the fluorescence traces that passed the above tests were used for the final localization analysis. The fluorescence images corresponding to these regions were subjected to a final fit using a pixelated Gaussian function to determine the centroid position. Because the CCD chip included square pixels of finite size, for optimum accuracy, the image was fit to:

$$I(x,y) = A + \int_{x-\delta}^{x+\delta} dX \int_{y-\delta}^{y+\delta} dY \, I_0 e^{\left[-\left(\frac{X-x_0}{a}\right)^2 - \left(\frac{Y-y_0}{b}\right)^2\right]/2}$$

which, for ease of evaluation, can be re-expressed in terms of error functions as:

$$I(x,y) = A + I_0 \frac{ab\pi}{4}\left[\text{erf}\left(\frac{x+\delta-x_0}{a}\right) - \text{erf}\left(\frac{x-\delta-x_0}{a}\right)\right]\left[\text{erf}\left(\frac{y+\delta-y_0}{b}\right) - \text{erf}\left(\frac{y-\delta-y_0}{b}\right)\right]$$

where A, I_o, a, b, x_o and y_o are as defined previously and δ (delta) is the fixed half-width of a pixel in the object plane. The final centroid coordinates (x_o, y_o) obtained from this fit were used as one data point in the final STORM image.

In another example of image analysis, an image movie was prepared that included a repetitive sequence of activation frames (in which the activation laser is on) and imaging frames (in which the imaging laser is on). For each imaging frame, fluorescent spots were identified and fit to Gaussian and/or elliptical Gaussian functions

to determine their centroid positions, intensities, widths, and ellipticities. Based on these parameters, peaks too dim, too wide or too skewed to yield satisfactory localization accuracy were rejected from further analysis. Peaks appearing in consecutive imaging frames with a displacement smaller than one camera pixel were considered to originate from the same fluorescent molecule and centroid positions of these peaks were connected across frames and organized into a data structure, which is referred to as a "string." Each string represents a single switching cycle for one fluorescent reporter molecule: the starting point of the string is the frame in which the molecule is switched on and its endpoint is the frame in which the molecule switches off. The final localization of the molecule was determined as the weighted average of the centroid positions across the entire string, weighted by the peak intensity of each frame. The total number of photons detected for each switching cycle was used as an additional filter to further reject localizations with low accuracy. Strings starting in an imaging frame immediately after an activation frame were recognized as a controlled activation event and color-coded according to the activation laser color. Other strings were identified as non-specific activations, most likely induced by the red imaging laser.

To correct for mechanical drift in the microscope during imaging, the same fitting algorithm was used to automatically track the motion of several fluorescent beads in the field of view. The beads served as fiducial marks and their positions were sampled during each imaging cycle. The averaged motion of the beads was subtracted from the coordinates obtained for each single switch position, yielding a drift-corrected reconstructed image.

As another drift correction method, the activator fluorophores was imaged during the activation frame and the correlation function was calculated between the first activation frame and all subsequent activation frames. By tracking the centroid of the correlation function, the drift of the image could be determined and corrected for in the STORM image. The correlation functions obtained from the fiducial marker images may also be used for drift correction. In some cases, it was also found that some further drift correction was possible by analyzing the correlation function of the STORM image itself as a function of time.

For quantitative analysis of images using switches equally spaced on dsDNA, the coordinates in the reconstructed image were classified using a A-means clustering

algorithm, and inter-switch distances were calculated as the distance between the cluster centroids. The number of clusters input to the algorithm was chosen according to the number of photobleaching steps observed during the initial exposure to red light.

Prediction of DNA configuration. Possible configurations of a 135 bp piece of dsDNA bound to the surface were computed by Monte Carlo simulation of the DNA as a worm-like chain in the plane. According to the Watson-Crick structure of dsDNA and accounting the length of the C_6 linkers attaching the Cy5 molecules to the nucleotides, the expected contour length between neighboring Cy5 molecules of 46 ± 1 nm was calculated. The DNA was treated as a series of 1 A (Angstrom) joints lying in the plane, each of which was deflected by a random angle selected from a Gaussian distribution chosen to give a persistence length of 50 nm. The measured inter-switch spacing was compared to the distribution of end-to-end distances of the polymer in this simulation.

While several embodiments of the present invention have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the functions and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the present invention. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the teachings of the present invention is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, the invention may be practiced otherwise than as specifically described and claimed. The present invention is directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles, materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the scope of the present invention.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

The indefinite articles "a" and "an," as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean "at least one."

The phrase "and/or," as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with "and/or" should be construed in the same fashion, i.e., "one or more" of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the "and/or" clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to "A and/or B", when used in conjunction with open-ended language such as "comprising" can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, "or" should be understood to have the same meaning as "and/or" as defined above. For example, when separating items in a list, "or" or "and/or" shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as "only one of or "exactly one of," or, when used in the claims, "consisting of," will refer to the inclusion of exactly one element of a number or list of elements. In general, the term "or" as used herein shall only be interpreted as indicating exclusive alternatives (i.e. "one or the other but not both") when preceded by terms of exclusivity, such as "either," "one of," "only one of," or "exactly one of." "Consisting essentially of," when used in the claims, shall have its ordinary meaning as used in the field of patent law.

As used herein in the specification and in the claims, the phrase "at least one," in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not

necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase "at least one" refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, "at least one of A and B" (or, equivalently, "at least one of A or B," or, equivalently "at least one of A and/or B") can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

In the claims, as well as in the specification above, all transitional phrases such as "comprising," "including," "carrying," "having," "containing," "involving," "holding," "composed of," and the like are to be understood to be open-ended, i.e., to mean including but not limited to. Only the transitional phrases "consisting of and "consisting essentially of shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

What is claimed is:

CLAIMS

1. A method, comprising:

providing a first entity and a second entity separated by a distance of less than about 1000 nm;

determining light emitted by the first entity;

determining light emitted by the second entity; and

determining the positions of the first entity and the second entity by using the light emitted by the first entity and the light emitted by the second entity.

2. The method of claim 1, wherein the light emitted by the first entity and the light emitted by the second entity have different wavelengths.

3. The method of claim 1, wherein the light emitted by the first entity and the light emitted by the second entity have substantially the same wavelength.

4. The method of claim 1, wherein the first entity and the second entity are activated by different wavelengths.

5. The method of claim 1, wherein the first entity and the second entity are activated by substantially the same wavelength.

6. The method of claim 1, comprising activating the first entity to produce the light emitted by the first entity.

7. The method of claim 6, wherein the first entity is activated by exposing the first entity to incident light of a first wavelength.

8. The method of claim 6, comprising deactivating the first entity to not produce the light emitted by the first entity.

9. The method of claim 8, wherein the first entity is deactivated by exposing the first entity to incident light of a second wavelength.

10. The method of claim 8, wherein the first entity is deactivated by exposing the first entity to incident light of the first wavelength.

11. The method of claim 1, further comprising activating the first entity to produce light but not the second entity.

12. The method of claim 11, further comprising activating the second entity to produce light.

13. The method of claim 12, further comprising deactivating the first entity.

14. The method of claim 13, wherein the act of activating the second entity to produce light occurs after the act of deactivating the first entity.

15. The method of claim 1, wherein the first entity is chemically different from the second entity.

16. The method of claim 1, wherein the first entity is substantially chemically identical to the second entity.

17. The method of claim 1, wherein the first entity is a photoactivatable or photoswitchable probe, and the second entity is a photoactivatable or photoswitchable probe.

18. The method of claim 1, wherein the first entity is a photoactivatable or photoswitchable dye.

19. The method of claim 1, wherein the first entity is a photoactivatable or photoswitchable fluorescent protein.

20. The method of claim 1, wherein the second entity is a photoactivatable or photoswitchable dye.

21. The method of claim 1, wherein the second entity is a photoactivatable or photoswitchable fluorescent protein.

22. The method of claim 1, wherein the first entity comprises a first, light-emitting portion and a second, activation portion that activates the first portion upon exposure to an external stimulus.

23. The method of claim 22, wherein the first, light-emitting portion is Cy5.

24. The method of claim 22, wherein the first, light-emitting portion is Cy5.5.

25. The method of claim 22, wherein the first, light-emitting portion is Cy7.

26. The method of claim 22, wherein the first, light-emitting portion is Alexa Fluor 647.

27. The method of claim 22, wherein the second, activation portion is Alexa Fluor 405.

28. The method of claim 22, wherein the second, activation portion is Alexa Fluor 488.

29. The method of claim 22, wherein the second, activation portion is Cy2.

30. The method of claim 22, wherein the second, activation portion is Cy3.

31. The method of claim 22, wherein the second, activation portion is Cy3.5.

32. The method of claim 22, wherein the second, activation portion is Cy5.

33. The method of claim 22, wherein the second entity comprises a first, light-emitting portion and a second, activation portion that activates the first portion upon exposure to an external stimulus.

34. The method of claim 33, wherein the first, light-emitting portion is Cy5.

35. The method of claim 33, wherein the first, light-emitting portion is Cy5.5.

36. The method of claim 33, wherein the first, light-emitting portion is Cy7.

37. The method of claim 33, wherein the first, light-emitting portion is Alexa Fluor 647.

38. The method of claim 33, wherein the second, activation portion is Alexa Fluor 405.

39. The method of claim 33, wherein the second, activation portion is Alexa Fluor 488.

40. The method of claim 33, wherein the second, activation portion is Cy2.

41. The method of claim 33, wherein the second, activation portion is Cy3.

42. The method of claim 33, wherein the second, activation portion is Cy3.5.

43. The method of claim 33, wherein the second, activation portion is Cy5.

44. The method of claim 33, wherein the first portion of the first entity and the first portion of the second entity are substantially chemically identical, and the second

portion of the first entity and the second portion of the second entity are not chemically identical.

45. The method of claim 33, wherein the first portion of the first entity and the first portion of the second entity are not chemically identical, and the second portion of the first entity and the second portion of the second entity are substantially chemically identical.

46. The method of claim 33, wherein the first portion of the first entity and the first portion of the second entity are not chemically identical, and the second portion of the first entity and the second portion of the second entity are not chemically identical.

47. The method of claim 33, wherein the first portion of the first entity and the first portion of the second entity are substantially chemically identical, and the second portion of the first entity and the second portion of the second entity are substantially chemically identical.

48. The method of claim 1, wherein the first entity and the second entity are switchable.

49. The method of claim 1, wherein the light emitted by the first entity is visible light.

50. The method of claim 1, wherein the first entity and the second entity are immobilized relative to a common entity.

51. The method of claim 50, wherein the common entity comprises a biomolecular complex.

52. The method of claim 50, wherein the common entity comprises a biological molecule.

53. The method of claim 52, wherein the biological molecule is a nucleic acid.

54. The method of claim 53, further comprising determining the number of bases separating the first entity and the second entity.

55. The method of claim 53, wherein the biological molecule is DNA.

56. The method of claim 53, wherein the biological molecule is RNA.

57. The method of claim 53, wherein the biological molecule is PNA.

58. The method of claim 52, wherein the biological molecule is a protein.

59. The method of claim 50, wherein the common entity is a cell.

60. The method of claim 50, wherein the common entity is a tissue.

61. The method of claim 50, wherein the common entity is nonbiological.

62. The method of claim 1, wherein the acts of determining light emitted by the first entity and determining light emitted by the second entity are repeated.

63. The method of claim 1, wherein the act of determining light emitted by the first entity comprises acquiring an image of the light emitted by the first entity.

64. The method of claim 1, wherein the act of determining the positions of the first entity and the second entity comprises using Gaussian fitting of the emitted light to determine the positions of the first entity and the second entity.

65. The method of claim 64, wherein the act of determining the positions of the first entity and the second entity further comprises using drift correction to determine the positions of the first entity and the second entity.

66. The method of claim 65, wherein the drift correction is performed by using fiduciary markers.

67. The method of claim 65, wherein the drift correction is performed by using image correlations.

68. The method of claim 1, wherein the positions of the first entity and the second entity are determined to an accuracy of at least about 300 niti.

69. The method of claim 1, wherein the positions of the first entity and the second entity are determined to an accuracy of at least about 100 nm.

70. The method of claim 1, wherein the positions of the first entity and the second entity are determined to an accuracy of at least about 50 nm.

71. The method of claim 1, wherein the positions of the first entity and the second entity are determined to an accuracy of at least about 20 nm.

72. The method of claim 1, wherein the positions of the first entity and the second entity are determined to an accuracy that is less than the wavelength of the light emitted by the first entity and/or the second entity.

73. The method of claim 1, wherein the first entity and a second entity separated by a distance of less than about 700 nm.

74. The method of claim 1, wherein the first entity and a second entity separated by a distance of less than the wavelength of light emitted by the first entity or the wavelength of light emitted by the second entity.

75. The method of claim 1, comprising determining the positions of the first entity and the second entity at a first point of time, and determining the positions of the first entity and the second entity at a second point of time.

76. The method of claim 1, comprising determining the positions of the first entity and the second entity at more than one points of time and/or as a function of time.

77. The method of claim 1, wherein the acts are performed in the order recited.

78. A method, comprising:

providing a first entity and a second entity separated by a distance of less than about 1000 nm;

activating the first entity but not the second entity;

determining light emitted by the first entity;

activating the second entity;

determining light emitted by the second entity; and

determining the positions of the first entity and the second entity by using the light emitted by the first entity and the light emitted by the second entity.

79. A method, comprising:

providing a first entity and a second entity separated by a distance of separation;

determining light emitted by the first entity, the light emitted by the first entity having a wavelength greater than the distance of separation;

determining light emitted by the second entity; and

determining the positions of the first entity and the second entity by using the light emitted by the first entity and the light emitted by the second entity.

80. The method of claim 79, wherein the light emitted by the first entity and the light emitted by the second entity have different wavelengths.

81. The method of claim 79, wherein the light emitted by the first entity and the light emitted by the second entity have substantially the same wavelength.

82. The method of claim 79, wherein light used to activate the first entity and light used to activate the second entity have different wavelengths.

83. The method of claim 79, wherein light used to activate the first entity and light used to activate the second entity have substantially the same wavelength.

84. The method of claim 79, comprising activating the first entity to produce the light emitted by the first entity.

85. The method of claim 84, wherein the first entity is activated by exposing the first entity to incident light of a first wavelength.

86. The method of claim 84, comprising deactivating the first entity to not produce the light emitted by the first entity.

87. The method of claim 86, wherein the first entity is deactivated by exposing the first entity to incident light of a second wavelength.

88. The method of claim 86, wherein the first entity is deactivated by exposing the first entity to incident light of the first wavelength.

89. The method of claim 79, further comprising activating the first entity to produce light but not the second entity.

90. The method of claim 89, further comprising activating the second entity to produce light.

91. The method of claim 90, further comprising deactivating the first entity.

92. The method of claim 91, wherein the act of activating the second entity to produce light occurs after the act of deactivating the first entity.

93. The method of claim 79, wherein the first entity is chemically different from the second entity.

94. The method of claim 79, wherein the first entity is substantially chemically identical to the second entity.

95. The method of claim 79, wherein the first entity is a photoactivatable or photoswitchable probe, and the second entity is a photoactivatable or photoswitchable probe.

96. The method of claim 79, wherein the first entity comprises a first, light-emitting portion and a second, activation portion that activates the first portion upon exposure to an external stimulus.

97. The method of claim 96, wherein the first, light-emitting portion is Cy5.

98. The method of claim 96, wherein the first, light-emitting portion is Cy5.5.

99. The method of claim 96, wherein the first, light-emitting portion is Cy7.

100. The method of claim 96, wherein the first, light-emitting portion is Alexa Fluor 647.

101. The method of claim 96, wherein the second, activation portion is Alexa Fluor 405.

102. The method of claim 96, wherein the second, activation portion is Alexa Fluor 488.

103. The method of claim 96, wherein the second, activation portion is Cy2.

104. The method of claim 96, wherein the second, activation portion is Cy3.

105. The method of claim 96, wherein the second, activation portion is Cy3.5.

106. The method of claim 96, wherein the second, activation portion is Cy5.

107. The method of claim 79, wherein the second entity comprises a first, light-emitting portion and a second, activation portion that activates the first portion upon exposure to an external stimulus.

108. The method of claim 107, wherein the first portion of the first entity and the first portion of the second entity are substantially chemically identical, and the second portion of the first entity and the second portion of the second entity are not chemically identical.

109. The method of claim 107, wherein the first portion of the first entity and the first portion of the second entity are not chemically identical, and the second portion of the first entity and the second portion of the second entity are substantially chemically identical.

110. The method of claim 107, wherein the first portion of the first entity and the first portion of the second entity are not chemically identical, and the second portion of the first entity and the second portion of the second entity are not chemically identical.

111. The method of claim 107, wherein the first portion of the first entity and the first portion of the second entity are substantially chemically identical, and the second portion of the first entity and the second portion of the second entity are substantially chemically identical.

112. The method of claim 79, wherein the light emitted by the first entity is visible light.

113. The method of claim 79, wherein the first entity and the second entity are immobilized relative to a common entity.

114. The method of claim 79, wherein the acts of determining light emitted by the first entity and determining light emitted by the second entity are repeated.

115. The method of claim 79, wherein the act of determining light emitted by the first entity comprises acquiring an image of the light emitted by the first entity.

116. The method of claim 79, wherein the act of determining the positions of the first entity and the second entity comprises using Gaussian fitting of the emitted light to determine the positions of the first entity and the second entity.

117. The method of claim 116, wherein the act of determining the positions of the first entity and the second entity further comprises using drift correction to determine the positions of the first entity and the second entity.

118. The method of claim 117, wherein the drift correction is performed by using fiduciary markers.

119. The method of claim 117, wherein the drift correction is performed by using image correlations.

120. The method of claim 79, wherein the positions of the first entity and the second entity are determined to an accuracy of at least about 300 nm.

121. The method of claim 79, wherein the positions of the first entity and the second entity are determined to an accuracy greater than the wavelength of the light emitted by the first entity or the second entity.

122. The method of claim 79, comprising determining the positions of the first entity and the second entity at a first point of time, and determining the positions of the first entity and the second entity at a second point of time.

123. The method of claim 79, comprising determining the positions of the first entity and the second entity at more than one point of time and/or as a function of time.

124. A method, comprising:

providing a first entity and a second entity separated by a distance of separation;

activating the first entity but not the second entity;

determining light emitted by the first entity, the light emitted by the first entity having a wavelength greater than the distance of separation;

activating the second entity;

determining light emitted by the second entity; and

determining the positions of the first entity and the second entity by using the light emitted by the first entity and the light emitted by the second entity.

125. A method, comprising:

providing a plurality of entities able to emit light, at least some of which are separated by a distance of separation less than the wavelength of the emitted light;

activating a fraction of the plurality of entities to emit light;

determining the emitted light;

deactivating the activated fraction of the plurality of entities; and

repeating the acts of activating and deactivating the plurality of entities to determine the positions of the plurality of entities.

126. The method of claim 125, wherein at least some of the plurality of entities emit light at different wavelengths.

127. The method of claim 125, wherein the plurality of entities each emit light at substantially the same wavelength.

128. The method of claim 125, wherein at least some of the plurality of entities are activated by light at different wavelengths.

129. The method of claim 125, wherein the plurality of entities are activated by light at substantially the same wavelength.

130. The method of claim 125, wherein the fraction of the plurality of entities are activated by exposing the plurality of entities to incident light of a first wavelength.

131. The method of claim 125, wherein the fraction of the plurality of entities are deactivated by exposing the fraction of the plurality of entities to incident light of a second wavelength.

132. The method of claim 125, wherein the fraction of the plurality of entities are deactivated by exposing the plurality of entities to incident light of the first wavelength.

133. The method of claim 125, wherein at least some of the plurality of entities are chemically different.

134. The method of claim 125, wherein the plurality of entities are substantially chemically identical.

135. The method of claim 125, wherein at least some of the plurality of entities are photoactivatable or photoswitchable probes.

136. The method of claim 125, wherein at least some of the plurality of entities comprises a first, light-emitting portion and a second, activation portion that activates the first portion upon exposure to an external stimulus.

137. The method of claim 136, wherein the first portions of the entities are chemically identical, and the second portions of the entities are not chemically identical.

138. The method of claim 136, wherein the first portions of the entities are not chemically identical, and the second portions of the entities are substantially chemically identical.

139. The method of claim 136, wherein the first portions of the entities are not chemically identical, and the second portions of the entities are not chemically identical.

140. The method of claim 136, wherein the first portions of the entities are substantially chemically identical, and the second portions of the entities are substantially chemically identical.

141. The method of claim 125, wherein the emitted light is visible light.

142. The method of claim 125, wherein at least some of the plurality of entities are immobilized relative to a common entity.

143. The method of claim 125, wherein the act of determining the emitted light comprises acquiring an image of the emitted light.

144. The method of claim 125, wherein the act of determining the positions of the plurality of entities comprises using Gaussian fitting of the emitted light to determine the positions of the plurality of entities.

145. The method of claim 144, wherein the act of determining the positions of the plurality of entities comprises using drift correction to determine the positions of the plurality of entities.

146. The method of claim 145, wherein the drift correction is performed by using fiduciary markers.

147. The method of claim 145, wherein the drift correction is performed by using image correlations.

148. The method of claim 125, wherein the positions of the plurality of entities are determined to an accuracy of at least about 300 nm.

149. The method of claim 125, wherein the positions of the plurality of entities are determined to an accuracy greater than the wavelength of the light emitted by the entities.

150. The method of claim 125, comprising determining the positions of the plurality of entities at a first point of time, and determining the positions of the plurality of entities at a second point of time.

151. The method of claim 125, comprising determining the positions of the plurality of entities at more than one point of time and/or as a function of time.

152. The method of claim 125, wherein the acts are performed in the order recited.

153. A method, comprising :

providing a plurality of entities able to emit light, at least some of which are separated by a distance of less than about 1000 nm;

activating a fraction of the plurality of entities to emit light;

determining the emitted light;

deactivating the activated fraction of the plurality of entities; and

repeating the acts of activating and deactivating the plurality of entities to determine the positions of the plurality of entities.

154. The method of claim 153, wherein at least some of the plurality of entities emit light at different wavelengths.

155. The method of claim 153, wherein the plurality of entities each emit light at substantially the same wavelength.

156. The method of claim 153, wherein at least some of the plurality of entities are activated by light at different wavelengths.

157. The method of claim 153, wherein the plurality of entities are activated by light at substantially the same wavelength.

158. The method of claim 153, wherein the fraction of the plurality of entities are activated by exposing the fraction of the plurality of entities to incident light of a first wavelength.

159. The method of claim 153, wherein the fraction of the plurality of entities are deactivated by exposing the fraction of the plurality of entities to incident light of a second wavelength.

160. The method of claim 153, wherein the fraction of the plurality of entities are deactivated by exposing the fraction of the plurality of entities to incident light of the first wavelength.

161. The method of claim 153, wherein at least some of the plurality of entities are photoactivatable or photoswitchable probes.

162. The method of claim 161, wherein at least some of the probes are photoactivatable or photoswitchable dyes.

163. The method of claim 161, wherein at least some of the probes are photoactivatable or photoswitchable fluorescent proteins.

164. The method of claim 153, wherein at least some of the plurality of entities comprises a first, light-emitting portion and a second, activation portion that activates the first portion upon exposure to an external stimulus.

165. The method of claim 164, wherein the first portions of the entities are chemically identical, and the second portions of the entities are not chemically identical.

166. The method of claim 164, wherein the first portions of the entities are not chemically identical, and the second portions of the entities are substantially chemically identical.

167. The method of claim 164, wherein the first portions of the entities are not chemically identical, and the second portions of the entities are not chemically identical.

168. The method of claim 153, wherein at least some of the plurality of entities are immobilized relative to a common entity.

169. The method of claim 153, wherein the act of determining the emitted light comprises acquiring an image of the emitted light.

170. The method of claim 153, wherein the act of determining the positions of the first entity and the second entity comprises using Gaussian fitting of the emitted light to determine the positions of the first entity and the second entity.

171. The method of claim 170, wherein the act of determining the positions of the first entity and the second entity further comprises using drift correction to determine the positions of the first entity and the second entity.

172. The method of claim 171, wherein the drift correction is performed using fiduciary markers.

173. The method of claim 171, wherein the drift correction is performed using image correlations.

174. The method of claim 153, wherein the positions of the plurality of entities are determined to an accuracy of at least about 300 run.

175. The method of claim 153, wherein the positions of the plurality of entities are determined to an accuracy that is less than the wavelength of the light emitted by the first entity.

176. The method of claim 153, comprising determining the positions of the plurality of entities at a first point of time, and determining the positions of the plurality of entities at a second point of time.

177. A method, comprising:

providing a first entity and a second entity separated by a distance of less than about 1000 nm, the first entity and the second entity each immobilized relative to a common entity;

determining the positions of the first entity and the second entity at a first point of time;

determining the positions of the first entity and the second entity at a second point of time; and

determining movement of the common entity using the positions of the first and second entities at the first and second points of time.

178. A method, comprising:

providing a first entity and a second entity separated by a distance of separation, the first entity and the second entity each immobilized relative to a common entity;

determining the positions of the first entity and the second entity at a first point of time using light emitted by the first entity and light emitted by the second entity, the light emitted by the first entity having a wavelength greater than the distance of separation;

determining the positions of the first entity and the second entity at a second point of time; and

determining movement of the common entity using the positions of the first and second entities at the first and second points of time.

179. A method, comprising:

identifying, within a series of images in time, one or more light-emission regions, each generated by a single entity;

for each light-emission region, identifying the center of the light-emission region; and

for each light-emission region, reconstructing the position of the single entity generating the light-emission region at a resolution greater than the wavelength of the light emitted by the single entity.

180. The method of claim 179, wherein the act of identifying the center of the light-emission region comprises using least-squares fitting to a Gaussian function.

181. The method of claim 179, wherein the act of determining the positions of the first entity and the second entity further comprises using drift correction to determine the positions of the first entity and the second entity.

182. The method of claim 181, wherein the drift correction is performed by using fiduciary markers.

183. The method of claim 181, wherein the drift correction is performed by using image correlations.

184. The method of claim 179, wherein the single entity is a photoactivatable or photoswitchable probe.

185. The method of claim 179, wherein the single entity comprises a first, light-emitting portion and a second, activation portion that activates the first portion upon exposure to an external stimulus.

186. The method of claim 179, further comprising producing an image using the reconstructed positions of the one or more light-emission regions.

187. An article, comprising:

a machine-readable medium comprising a program, embodied in the medium, for causing a machine to perform a method comprising acts of:

identifying, within a series of images in time, one or more light-emission regions, each generated by a single entity;

for each light-emission region, identifying the center of the light-emission region; and

for each light-emission region, reconstructing the position of the single entity generating the light-emission region at a resolution greater than the wavelength of the light emitted by the single entity.

188. An article, comprising :

a translation stage for a microscope having a drift of less than about 100 nm/min.

189. The article of claim 188, wherein the translation stage has a drift of less than about 10 nm/min.

190. The article of claim 188, further comprising a light source positioned to illuminate at least a portion of the translation stage, the light source being activatable and deactivatable in a programmed fashion.

191. The article of claim 188, further comprising a light source positioned to illuminate at least a portion of the translation stage, the light source being activatable and deactivatable in a periodic fashion.

192. The article of claim 189, further comprising a light source positioned to illuminate at least a portion of the translation stage, the light source being modulatable using a shutter.

193. The article of claim 189, further comprising a light source positioned to illuminate at least a portion of the translation stage, the light source being modulatable using an acoustic optical modulator.

194. The article of claim 189, further comprising a light detector position in optical communication with the translation stage.

195. The article of claim 194, wherein the light detector comprises a photodiode, a photomultiplier, or a CCD camera.

196. The article of claim 189, further comprising more than one light source positioned to illuminate at least a portion of the translation stage.

197. The article of claim 196, wherein at least two of the more than one light source emit light at different wavelengths.

198. The article of claim 189, comprising a dichroic mirror or a polychroic mirror positioned to direct light from a light source to the translation stage.

199. The article of claim 189, further comprising a light source positioned to illuminate at least a portion of the translation stage, and a color filter positioned to alter at least a portion of the light emanating from the light source prior to illumination of the translation stage.

200. An imaging composition, comprising:

a light-emitting entity, capable of being switched between a first state able to emit light at a first, emission wavelength and a second state that does not substantially emit light at the first wavelength.

201. The imaging composition of claim 200, further comprising a second light-emitting entity, distinguishable from the light-emitting entity, capable of being switched between a first state able to emit light at a first, emission wavelength and a second state that does not substantially emit light at the first wavelength.

202. The imaging composition of claim 200, wherein the light-emitting entity comprises:

a first portion that is capable of emitting light at the first wavelength; and

a second portion that activates the first portion upon exposure to an external stimulus, thereby causing the first portion to emit light at the first wavelength.

203. The imaging composition of claim 202, wherein the second portion that activates the first portion upon exposure to light having an excitation wavelength.

204. The imaging composition of claim 202, wherein the first portion is covalently bonded to the second portion.

205. The imaging composition of claim 202, wherein the first portion and the second portion are each bonded to a common entity.

206. The imaging composition of claim 202, wherein the first portion, when isolated from the second portion, is fluorescent.

207. The imaging composition of claim 202, wherein the second portion, when isolated from the first portion, is fluorescent.

208. The imaging composition of claim 202, wherein the first portion is Cy5, Cy5.5, Cy7, or Alexa Fluor 647.

209. The imaging composition of claim 202, wherein the second portion is Alexa Fluor 405, Cy2, Alexa Fluor 488, Cy3, or Cy3.5.

210. The imaging composition of claim 202, wherein the first portion and the second portion are separated by a distance of no more than about 50 nm.

211. The imaging composition of claim 202, further comprising a second light-emitting entity, distinguishable from the light-emitting entity, comprising a first portion that is capable of emitting light at a first wavelength, and a second portion that activates the first portion upon exposure to an external stimulus, thereby causing the first portion to emit light at the first wavelength.

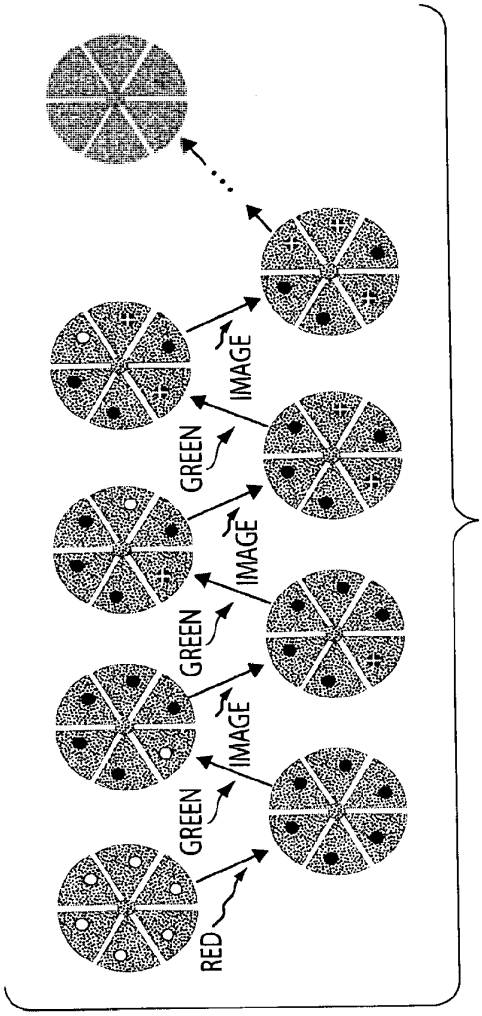


Fig. 1A

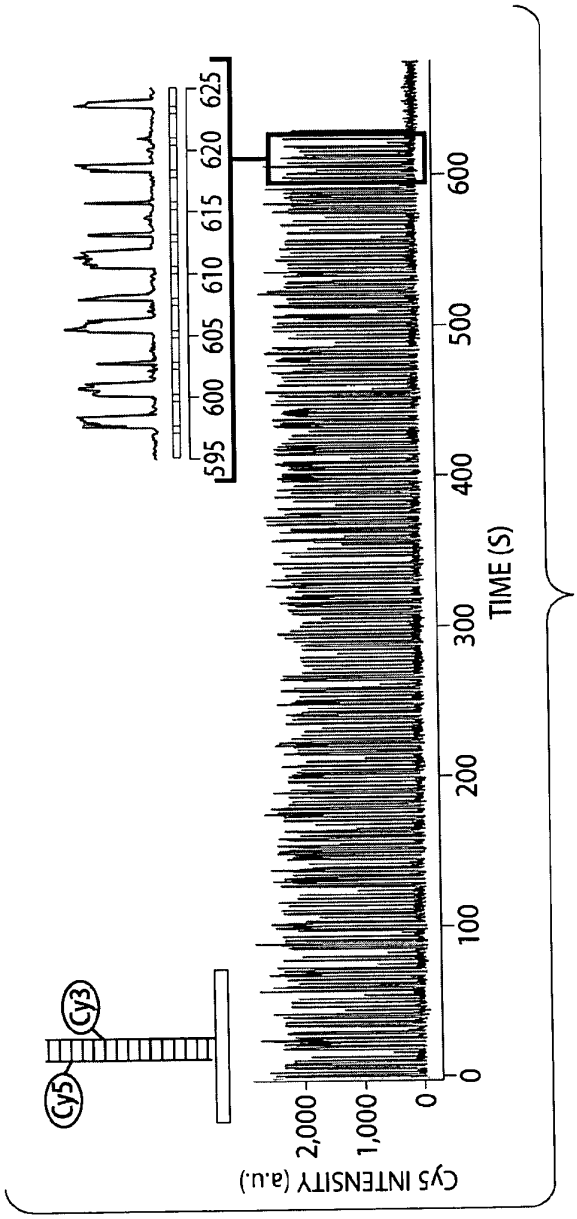

Cy5 INTENSITY (a.u.)
2,000
1,000
0
0
100
200
300
400
500
600
TIME (S)
595
600
605
610
615
620
625
Cy3
Cy5

Fig. 1B

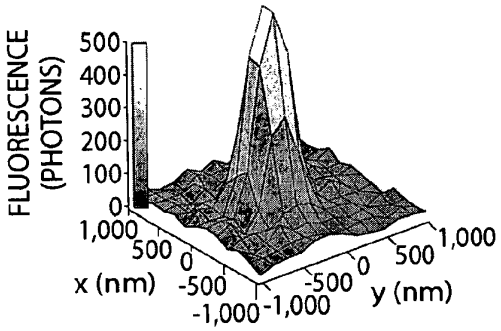


Fig. 2A

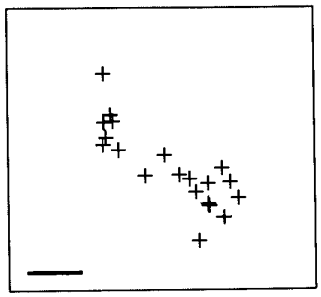


Fig. 2B

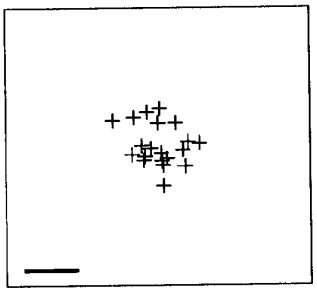


Fig. 2C

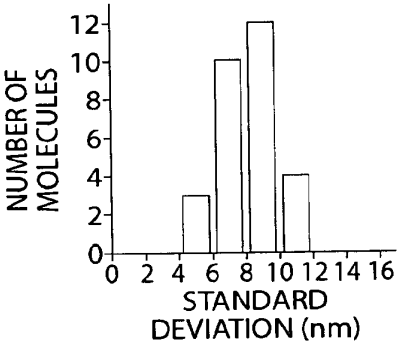


Fig. 2D

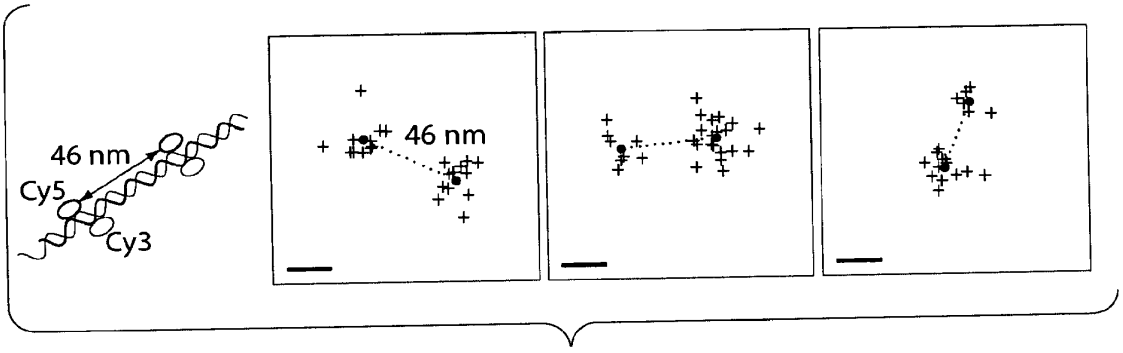

46 nm
Cy5
Cy3
46 nm

Fig. 3A

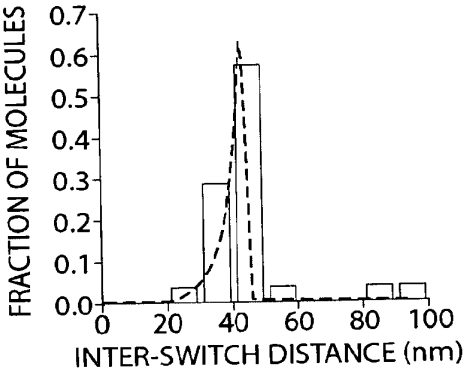

FRACTION OF MOLECULES
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
20
40
60
80
100
INTER-SWITCH DISTANCE (nm)

Fig. 3B

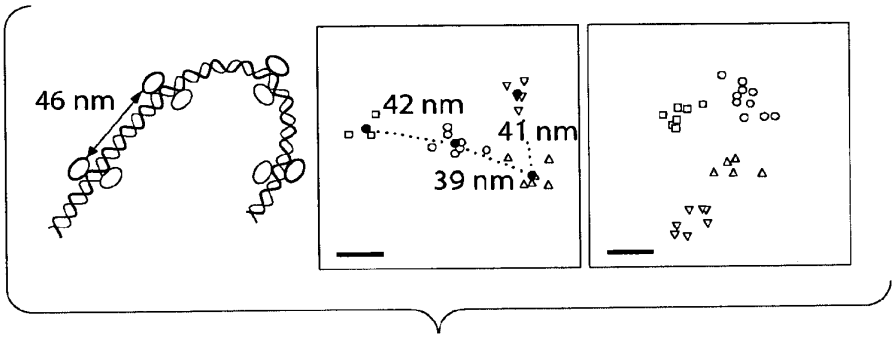


Fig. 3C

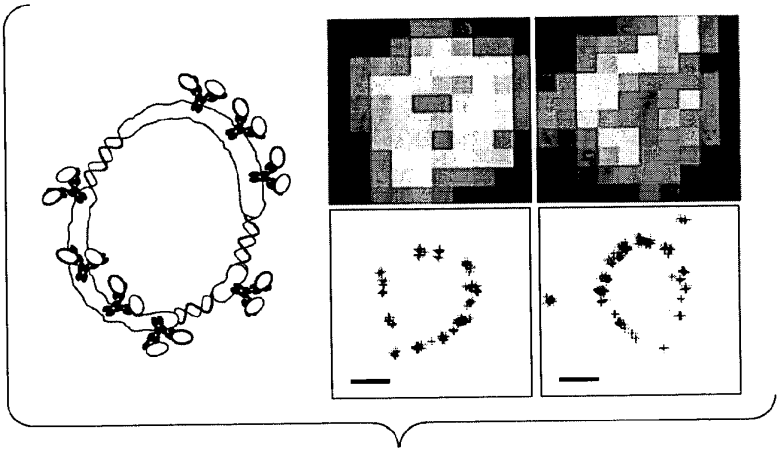


Fig. 3D

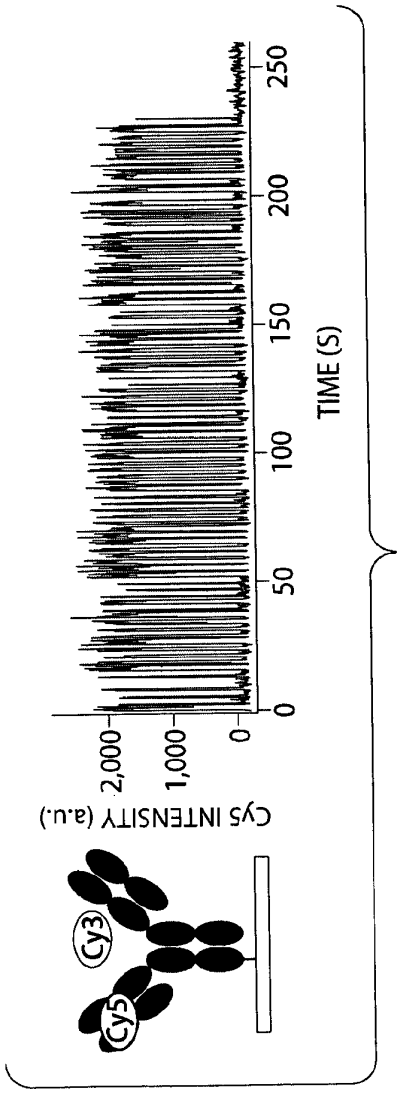


Fig. 4

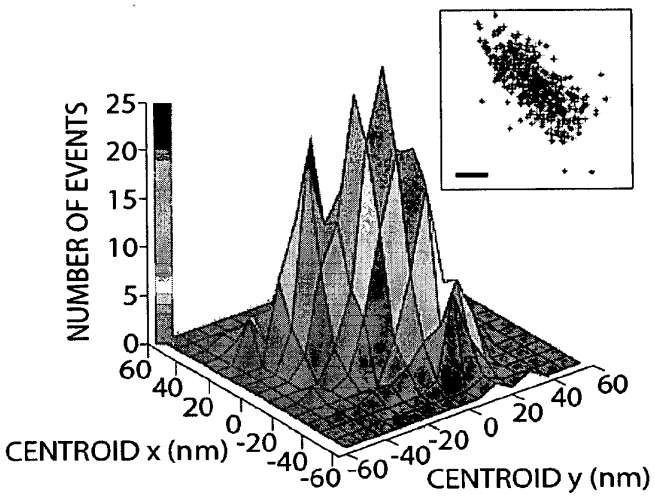

NUMBER OF EVENTS
25
20
15
10
5
0
60
40
20
0
-20
-40
-60
CENTROID x (nm)
-60
-40
-20
0
20
40
60
CENTROID y (nm)

Fig. 5A

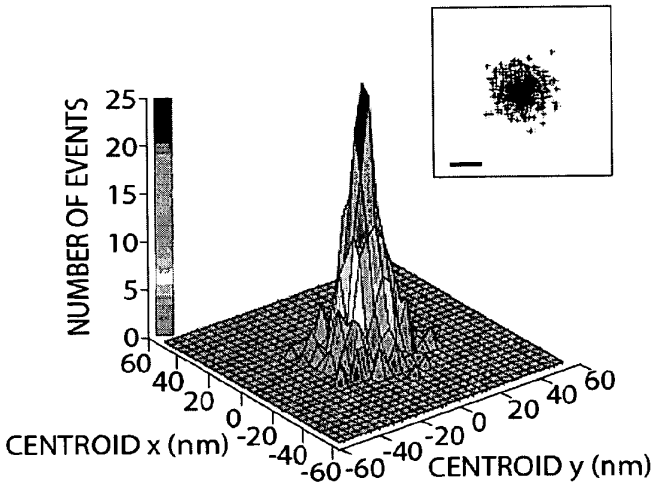

NUMBER OF EVENTS
25
20
15
10
5
0
60
40
20
0
-20
-40
-60
CENTROID x (nm)
-60
-40
-20
0
20
40
60
CENTROID y (nm)

Fig. 5B

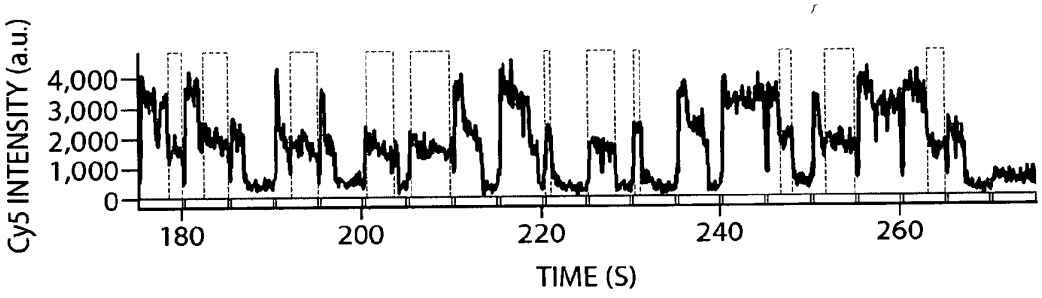

Cy5 INTENSITY (a.u.)
4,000
3,000
2,000
1,000
0
180
200
220
240
260
TIME (S)

Fig. 6

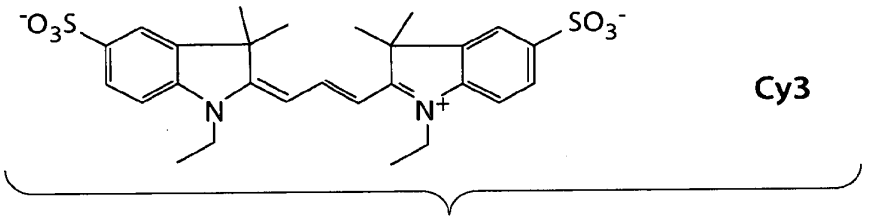


Fig. 7A

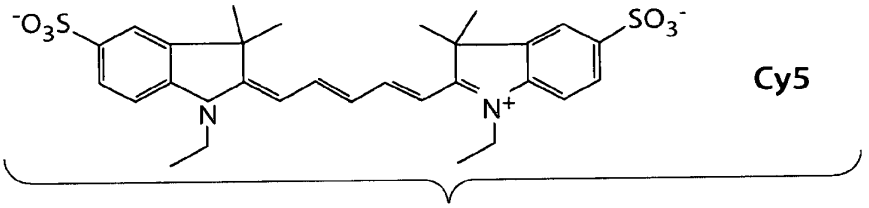


Fig. 7B

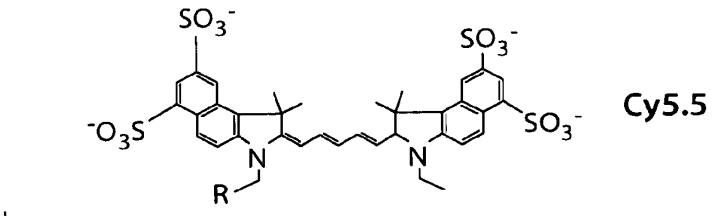


Fig. 7C

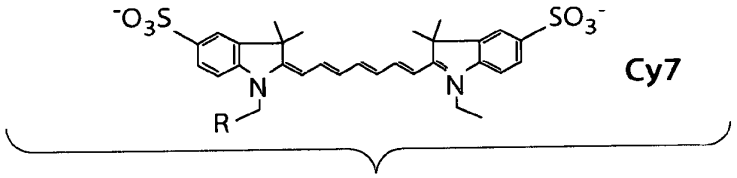


Fig. 7D

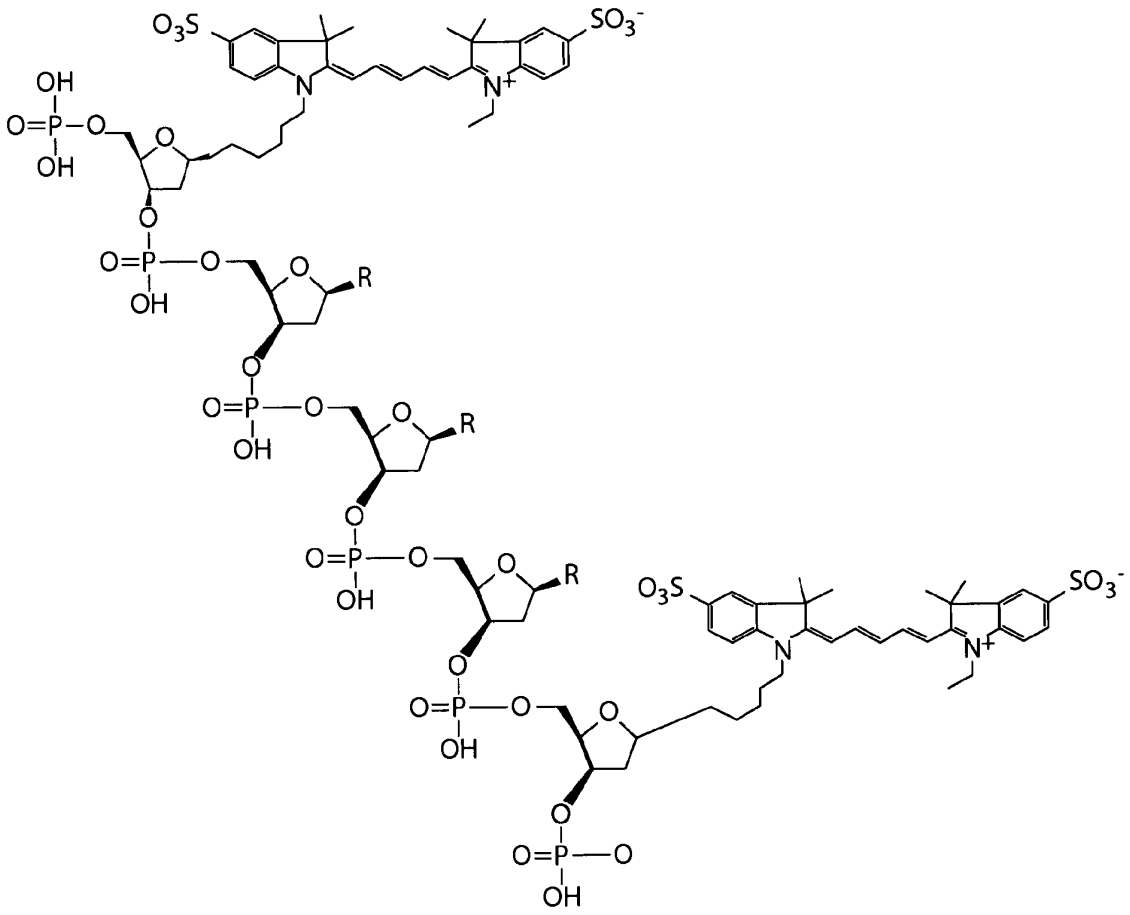


Fig. 7E

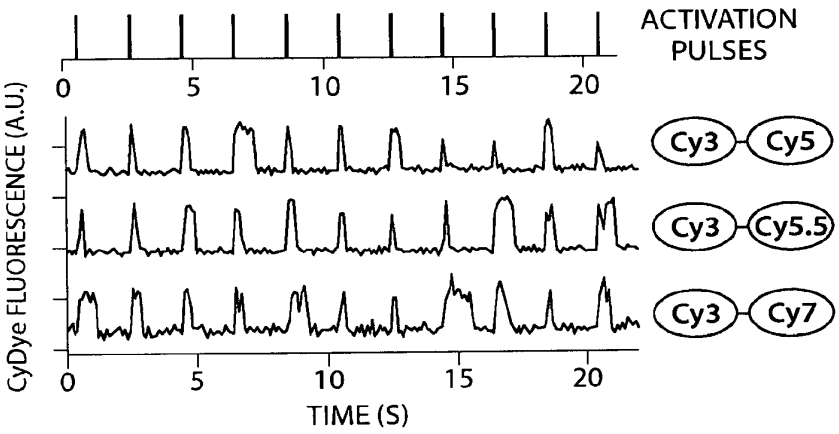

ACTIVATION PULSES
0
5
10
15
20
CyDye FLUORESCENCE (A.U.)
Cy3 Cy5
Cy3 Cy5.5
Cy3 Cy7
0
5
10
15
20
TIME (S)

Fig. 8A

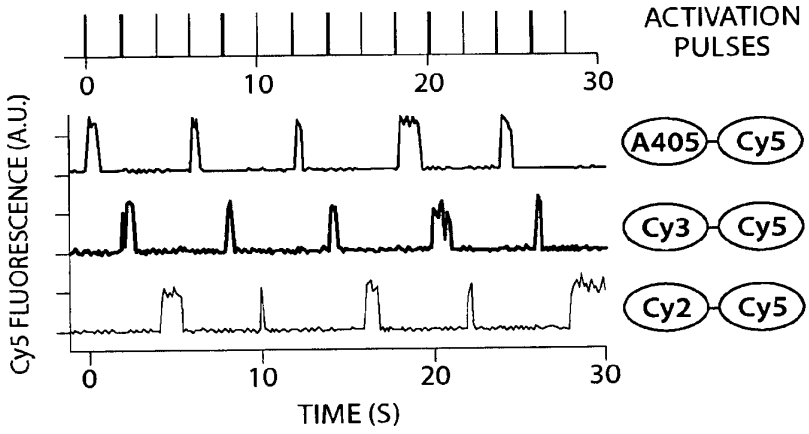

ACTIVATION PULSES
0
10
20
30
Cy5 FLUORESCENCE (A.U.)
A405 Cy5
Cy3 Cy5
Cy2 Cy5
0
10
20
30
TIME (S)

Fig. 8B

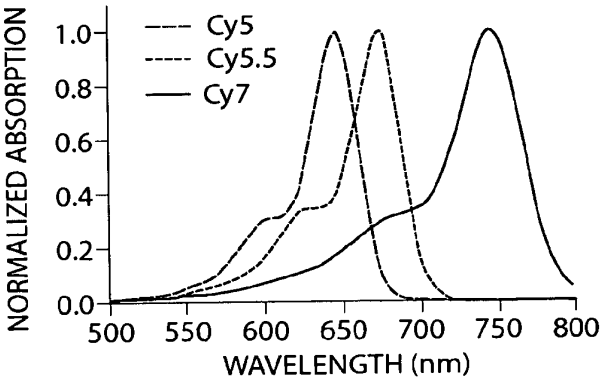


Fig. 9A

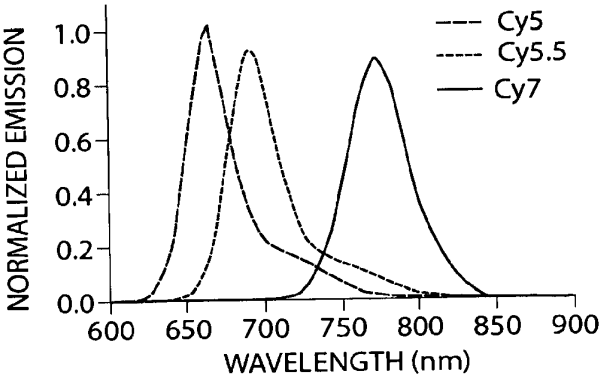


Fig. 9B

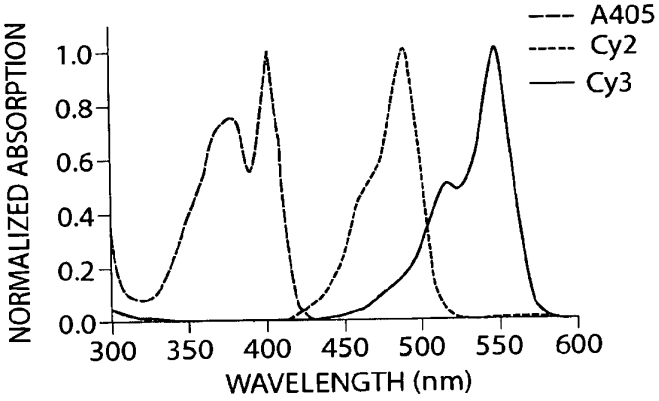


Fig. 9C

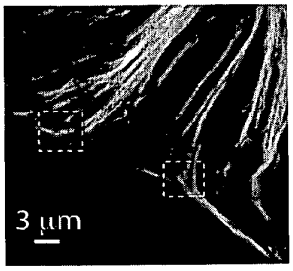


Fig. 10A

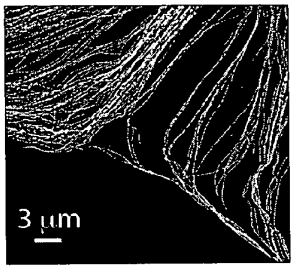


Fig. 10B

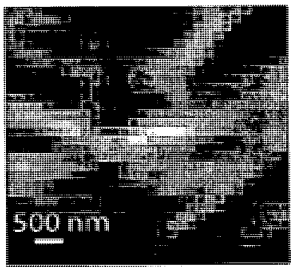


Fig. 10C

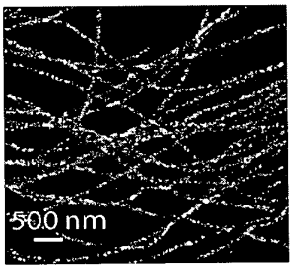


Fig. 10D

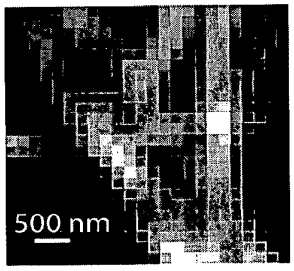


Fig. 10E

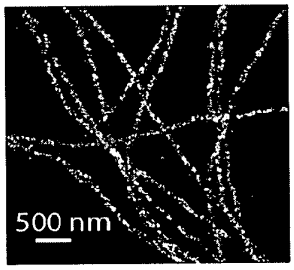


Fig. 10F

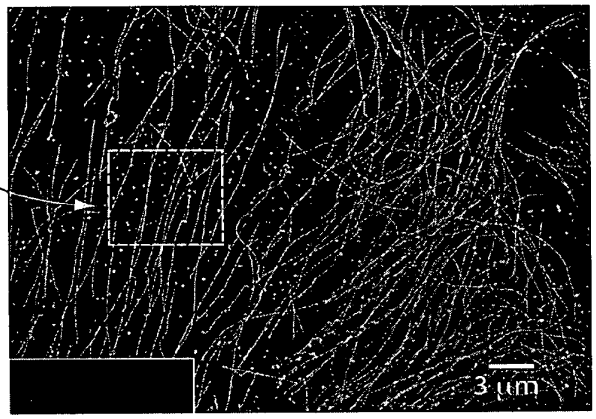


Fig. 11A

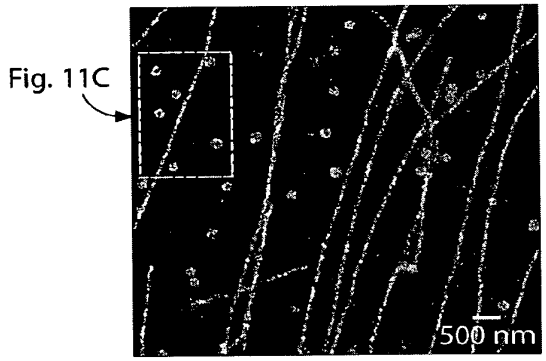


Fig. 11B

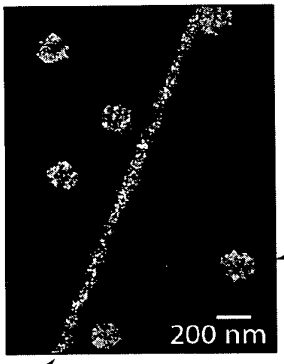


Fig. 11C

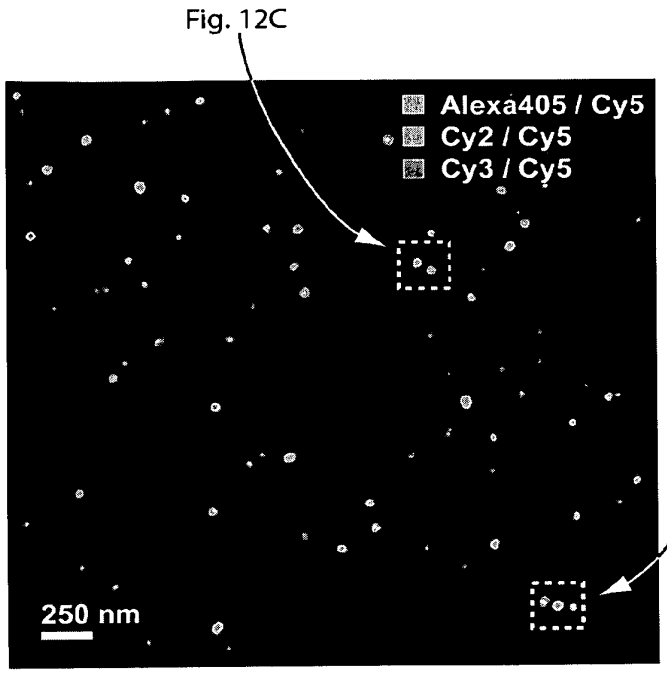


Fig.12A

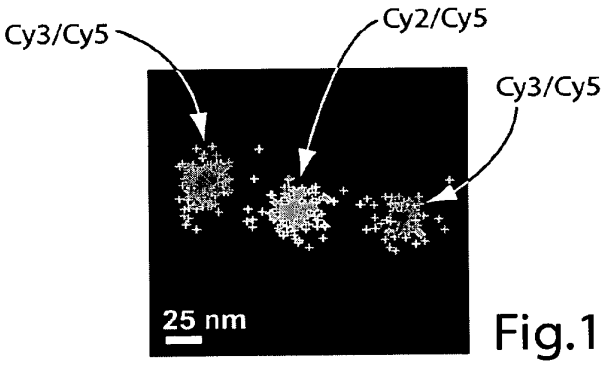


Fig.12B

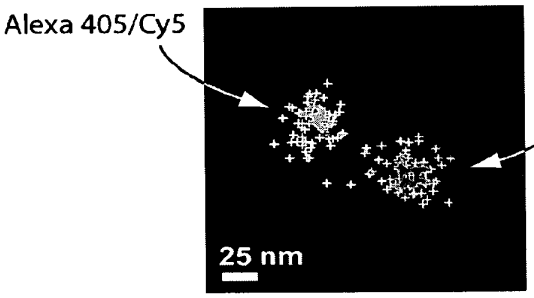


Fig.12C

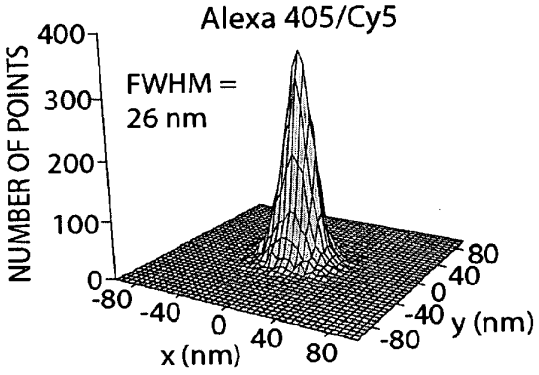


Fig. 12D

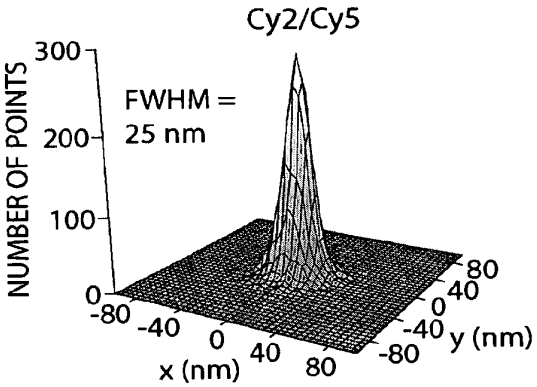


Fig. 12E

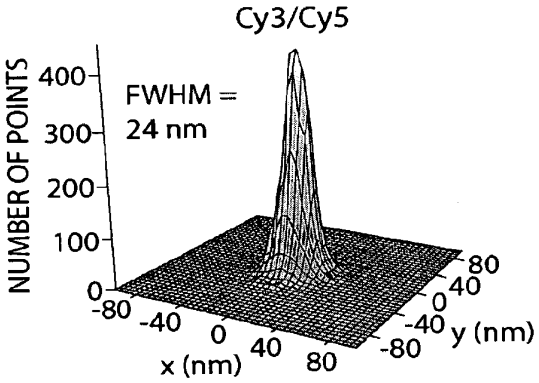


Fig. 12F

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 February 2016 (04.02.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/018963 A9

(51) International Patent Classification:
C12P 19/34 (2006.01) *C07H 21/04* (2006.01)
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2015/042559

(22) International Filing Date:
29 July 2015 (29.07.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/03 1,062 30 July 2014 (30.07.2014) US
62/050,636 15 September 2014 (15.09.2014) US
62/142,653 3 April 2015 (03.04.2015) US

(71) Applicant: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).

(72) Inventors: **ZHUANG, Xiaowei;** 5 Luongo Farm Lane, Lexington, MA 02421 (US). **MOFFITT, Jeffrey, R.;** 175 Beacon Street, Apt. #108, Somerville, MA 02143 (US). **BOETTIGER, Alistair;** 1550 Cambridge Street, Apt. #2, Cambridge, MA 02139 (US).

(74) Agent: **CHEN, Tani;** Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available):* AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

[Continued on nextpage]

(54) Title: PROBE LIBRARY CONSTRUCTION

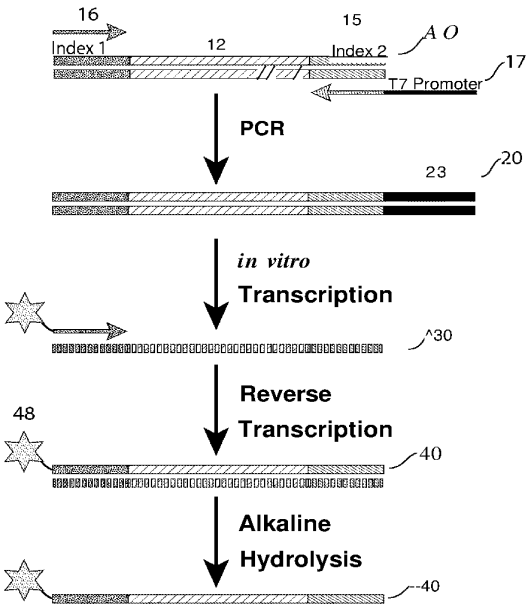


FIG. 1

(57) Abstract: The present invention generally relates to systems and methods for producing nucleic acids. In some aspects, relatively large quantities of oligonucleotides can be produced, and in some cases, the oligonucleotides may have a variety of different sequences and/or lengths. For instance, a relatively small quantity of oligonucleotides may be amplified to produce a large amount of nucleotides. In one set of embodiments, oligonucleotides may be amplified using PCR, then transcribed to produce RNA. The RNA may then be reverse transcribed to produce DNA, and optionally, the RNA may be selectively degraded or removed, relative to the DNA. In one set of embodiments, the oligonucleotides may be chemically modified. These modifications may include, but are not limited, to the adding of fluorescent dyes or other signaling entities.

WO 2016/018963 A9

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available):* ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

(48) **Date of publication of this corrected version:**
24 March 2016

(15) **Information about Correction:**
see Notice of 24 March 2016

PROBE LIBRARY CONSTRUCTION

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Patent Application Serial No. 62/050,636, filed September 15, 2014, entitled "Probe Library Construction," by Zhuang, *et al.;* U.S. Provisional Patent Application Serial No. 62/031,062, filed July 30, 2014, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.;* and U.S. Provisional Patent Application Serial No. 62/142,653, filed April 3, 2015, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.* Each of the above is incorporated herein by reference.

GOVERNMENT FUNDING

This invention was made with government support under Grant No. GM096450 awarded by the National Institutes of Health. The government has certain rights in the invention.

FIELD

The present invention generally relates to systems and methods for producing nucleic acids.

BACKGROUND

Custom-synthesized, oligonucleotide probes have emerged as a powerful tool for the identification and isolation of specific nucleic acid targets via hybridization. Applications for such hybridization probe sets range from next generation sequencing—where such probes are used to enrich or deplete samples for specific nucleic acid targets—to imaging of fixed samples—where fluorescently labeled hybridization probes allow the direct measurement of the number and spatial organization of the targeted species.

There are now a wide range of commercial sources for such probes. Such probes are often made by synthesizing each oligonucleotide member using standard solid phase synthesis methods. Unfortunately, this limits both the number of probes within a single set and the number of unique sets, due to the requirement that each oligonucleotide member must be individually and separately synthesized.

Recent advances in array-based synthesis of oligonucleotides by several companies have reduced the cost of producing oligonucleotides. However, these approaches also

result in 1000-fold less oligonucleotide probes than is required for a single hybridization reaction, thus limiting their usefulness. Accordingly, improvements in oligonucleotide production are needed.

SUMMARY

The present invention generally relates to systems and methods for producing nucleic acids. The subject matter of the present invention involves, in some cases, interrelated products, alternative solutions to a particular problem, and/or a plurality of different uses of one or more systems and/or articles.

In one aspect, the present invention is generally directed to a method. The method, in accordance with one set of embodiments, includes amplifying at least some of a plurality of oligonucleotides using real-time PCR to produce amplified oligonucleotides, transcribing *in vitro* at least some of the amplified oligonucleotides to produce RNA, reverse transcribing the RNA to produce transcribed DNA, and selectively degrading the RNA relative to the transcribed DNA.

In another set of embodiments, the method includes simultaneously amplifying at least some of a plurality of oligonucleotides in a common solution using PCR to produce amplified oligonucleotides, transcribing *in vitro* at least some of the amplified oligonucleotides to produce RNA, reverse transcribing the RNA to produce transcribed DNA, and selectively degrading the RNA relative to the transcribed DNA.

In yet another set of embodiments, the method includes acts of providing a plurality of oligonucleotides having an average length of between 10 and 200 nucleotides and including at least 100 unique oligonucleotide sequences, producing amplified oligonucleotides comprising one of the plurality of oligonucleotides and a promoter using real-time PCR, transcribing at least some of the amplified oligonucleotides to produce RNA using an RNA polymerase, reverse transcribing the RNA to produce DNA using a primer comprising a signaling entity, and chemically reducing the RNA.

The method, in still another set of embodiments, includes acts of providing a plurality of oligonucleotides having an average length of between 10 and 200 nucleotides and including at least 100 unique oligonucleotide sequences, producing amplified oligonucleotides in a common solution comprising one of the plurality of oligonucleotides and a promoter using PCR, transcribing at least some of the amplified oligonucleotides to produce RNA using an RNA polymerase, reverse transcribing the

RNA to produce DNA using a primer comprising a signaling entity, and chemically reducing the RNA.

In another aspect, the present invention encompasses methods of making one or more of the embodiments described herein, such as oligonucleotides, including but not limited to modified oligonucleotides such as those described herein (e.g., labeled with a signaling entity). In still another aspect, the present invention encompasses methods of using one or more of the embodiments described herein, for example, such as oligonucleotides, including but not limited to modified oligonucleotides such as those described herein (e.g., labeled with a signaling entity).

Other advantages and novel features of the present invention will become apparent from the following detailed description of various non-limiting embodiments of the invention when considered in conjunction with the accompanying figures. In cases where the present specification and a document incorporated by reference include conflicting and/or inconsistent disclosure, the present specification shall control. If two or more documents incorporated by reference include conflicting and/or inconsistent disclosure with respect to each other, then the document having the later effective date shall control.

BRIEF DESCRIPTION OF THE DRAWINGS

Non-limiting embodiments of the present invention will be described by way of example with reference to the accompanying figures, which are schematic and are not intended to be drawn to scale. In the figures, each identical or nearly identical component illustrated is typically represented by a single numeral. For purposes of clarity, not every component is labeled in every figure, nor is every component of each embodiment of the invention shown where illustration is not necessary to allow those of ordinary skill in the art to understand the invention. In the figures:

Fig. 1 illustrate the production of DNA probes, in accordance with one set of embodiments; and

Fig. 2 illustrates a template molecule (SEQ ID NO: 3) produced in accordance with one embodiment of the invention;

Fig. 3 lists optimized primers from the *E. coli* transcriptome, in another embodiment of the invention (the sequences, from top to bottom and left to right, by page, correspond to SEQ ID NOs: 4-57, 58-111, 112-165, and 166-201); and

Fig. 4 shows various probes in accordance with yet another embodiment of the invention. The sequences in Fig. 4, from top to bottom and left to right, by page, correspond to SEQ ID NOs: 202-221, 222-243, 244-265, 266-287, 288-309, 310-331, 332-353, 354-375, 376-397, 398-419, 420-441, 442-463, 464-485, 486-507, 508-529, 530-551, 552-573, 574-595, 596-617, 618-639, 640-661, 662-683, 684-705, 706-727, 728-749, 750-771, 772-793, 794-815, 816-837, 838-859, 860-881, 882-903, 904-925, and 926-937.

DETAILED DESCRIPTION

The present invention generally relates to systems and methods for producing nucleic acids. In some aspects, relatively large quantities of oligonucleotides can be produced, and in some cases, the oligonucleotides may have a variety of different sequences and/or lengths. For instance, a relatively small quantity of oligonucleotides may be amplified to produce a large amount of nucleotides. In one set of embodiments, oligonucleotides may be amplified using PCR, then transcribed to produce RNA. The RNA may then be reverse transcribed to produce DNA, and optionally, the RNA may be selectively degraded or removed, relative to the DNA. In one set of embodiments, the oligonucleotides may be chemically modified. These modifications may include, but are not limited, to the adding of fluorescent dyes or other signaling entities.

U.S. Provisional Patent Application Serial No. 62/031,062, filed July 30, 2014, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.* is incorporated herein by reference in its entirety.

In one aspect, the present invention is generally directed to *in vitro* methods of amplifying a plurality of oligonucleotides. In some cases, relatively large numbers of unique oligonucelotides within a plurality of oligonucleotides may be amplified. For instance, a plurality of oligonucleotides to be amplified may include 10, 100, 1,000, or more unique sequences.

In addition, in some embodiments, the oligonucleotides may be amplified without selective amplification of some oligonucleotides over others, e.g., due to competitive effects. Although some drift may occur, it is desired that the relative ratios of the oligonucleotides within a plurality of oligonucleotides stay substantially the same after amplification, at least for some applications. However, in many amplification techniques, due to differences in binding or affinity of different oligonucleotides, some

oligonucleotides may be amplified to a greater degree than others, and thus, specific techniques need to be utilized to reduce or eliminate this problem, for example, by separately amplifying each of the oligonucleotides before combining them together to form the plurality of oligonucleotides. In contrast, as is discussed herein, in certain embodiments, a plurality of oligonucleotides can be amplified without causing substantial alterations or changes in the ratios of the oligonucleotides, without requiring separation of the oligonucleotides, separate growth of the oligonucleotides, or other cumbersome techniques.

Referring now to Fig. 1, one example of an embodiment of the invention is now illustrated. In this figure, a plurality of oligonucleotides 10 is provided. This may include 1, 10, 100, 1,000, 10,000, 100,000, or any other suitable number of unique oligonucleotide sequences. Of course, more than one copy of any particular unique oligonucleotide may also be present as well within the plurality of oligonucleotides. The unique oligonucleotides may have the same or different lengths. In some cases, the plurality of oligonucleotides have an overall average length (number average or arithmetic mean) of less than 200 nt (nucleotides), although longer average lengths are also possible in some embodiments.

The plurality of oligonucleotides 10 may initially be amplified, using PCR (polymerase chain reaction) or another suitable oligonucleotide amplification method, to produce a plurality of amplified oligonucleotides 20. In some cases, PCR may be used to generate thousands to millions of copies per oligonucleotide within the plurality of oligonucleotides. In some embodiments, the plurality of oligonucleotides may be amplified while still contained in a common solution, for instance, without requiring separation of the oligonucleotides prior to amplification, e.g., as is required in certain techniques such as emulsion PCR.

Within a common solution, while it is possible that different oligonucleotides of the plurality of oligonucleotides may be amplified at different rates (e.g., leading to non-uniform amplification of the plurality of oligonucleotides, and the potential loss of complexity or species within the plurality of oligonucleotides during amplification), in certain embodiments, this can be reduced or minimized through the use of various oligonucleotide structures and/or through the use of certain types of PCR techniques, as is discussed herein.

As an example, in one set of embodiments, the plurality of oligonucleotides may all be chosen to minimize competitive effects, e.g., as caused by differences in binding or affinity of the oligonucleotides to reagents within the common solution or the preferential enzymatic amplification of some sequence features. For example, in one set of embodiments, the plurality of oligonucleotides may be chosen to have similar lengths and/or sequences.

As is shown in Fig. 1, the plurality of oligonucleotides may each contain one or more index regions 15, 16 on one or both ends of the oligonucleotides that can be recognized by certain reagents. In some cases, the oligonucleotides of the plurality of oligonucleotides may have one or more index regions to which suitable primers can interact with in order to allow PCR or other amplification to occur. In some embodiments, these index regions can be used to selectively produce DNA probes only from subset of the plurality of oligonucleotides 10.

These index regions can also be used in some instances to add additional sequences to that of the plurality of oligonucleotides, e.g., as is shown in Fig. 1, oligonucleotide 11 may include an index region 15, to which a sequence 17 containing a T7 promoter can bind and be introduced into the amplified oligonucleotides (e.g., as region 23). Various sequences may thus be applied to the plurality of oligonucleotides that include a portion able to bind an index region. In some cases, if substantially all of the plurality of oligonucleotides contain similar or identical index regions, the relative affinities or binding to the index regions of the oligonucleotides by enzymes such as polymerases may be substantially similar or identical, which may allow for relatively uniform amplification to occur. The plurality of oligonucleotides may also contain other different regions that can be varied to produce a plurality of unique oligonucleotides, e.g., region 12. These regions can vary in terms of length and/or sequence, etc.

In some embodiments, the amount of amplification that occurs may be carefully controlled by monitoring the PCR amplification reaction, e.g., using techniques such as real-time PCR. This may occur, for example, using oligonucleotides having common index regions, substantially similar lengths and/or sequences, etc., including those previously discussed, or with other suitable pluralities of oligonucleotides. For instance, in some embodiments, the PCR reaction may be monitored by illuminating the solution containing the oligonucleotides with suitable light and determining the amount of

fluorescence that is present, which can be related to the DNA present within the sample. Techniques for monitoring PCR reactions, such as real-time PCR methodologies, are known to those of ordinary skill in the art. The PCR reaction can also be controlled, in some embodiments, by controlling the amount and/or concentration of nucleotides and/or cofactors that are present.

After amplification, the plurality of amplified oligonucleotides 20 may be transcribed to produce a plurality of RNAs 30, as is shown in the example of Fig. 1. This may be performed, for example, by exposing the amplified oligonucleotides to a suitable RNA polymerase, such as T7 RNA polymerase, that can transcribe the oligonucleotides to produce corresponding RNA. The amount of RNA production can be controlled in some embodiments by controlling the amount and/or concentration of nucleotides and/or cofactors that are present as well as the duration of the *in vitro* transcription reaction.

The plurality of RNAs 30, may then be used to produce additional amounts of DNA 40, e.g., by using reverse transcription. For example a suitable enzyme, such as reverse transcriptase, may be used to perform the reverse transcription. In some cases, primers may be used to facilitate transcription, and in some embodiments, the primers may also be used to attach additional entities to the DNA. For example, signaling entities may be attached, as is shown with signaling entity 48 in Fig. 1. Alternatively, additional nucleic acid sequences can also be attached, which can serve to recruit additional oligonucleotides via Watson-Crick base-pairing. The amount of DNA that is produced can be controlled, for example, by controlling the amount and/or concentration of nucleotides and/or cofactors that are present as well as the duration and temperature of the reverse transcription reaction.

In some embodiments, multiple copies of DNA may be produced from each RNA molecule. In addition, optionally, the RNA may then be removed or selectively degraded, relative to the DNA, for example, through alkaline hydrolysis, enzymatic digestion, or other techniques.

Accordingly, in certain aspects, the present invention is generally directed to systems and methods of amplifying a plurality of oligonucleotides. In one set of embodiments, relatively large quantities or masses of oligonucleotides can be produced as is discussed herein, e.g., at least about 10^{-3} pmol, at least about 10^{-2} pmol, at least about 10^{-1} pmol, at least about 10^{0} pmol, at least about 10^{1} pmol, at least about 10^{2} pmol,

at least about 10^3 pmol, etc. In addition, in some embodiments, the plurality of oligonucleotides may be substantially diverse. For example, the plurality of oligonucleotides may include at least about 10^1, at least about 102, at least about 10^3, at least about 10^4, at least about 10^5, or at least about 10^6 unique sequences of oligonucleotides, even after amplification to the amounts discussed above. (It should also be noted that a plurality or population of oligonucleotides may include more than one copy of a given unique oligonucleotide sequence.) In contrast, certain prior art techniques are able to amplify large numbers of unique oligonucleotides, but to only small quantities or masses (e.g., to amounts of around 10^{-3} pmol or less), or are able to produce large quantities or masses of oligonucleotides, but only for 1 or a few unique sequences (e.g., less than 10 sequences).

As discussed, in certain embodiments, a plurality of oligonucleotides, which may include a plurality of unique sequences of oligonucleotides such as those described above, may be amplified without substantial selective amplification of some oligonucleotide sequences over others, e.g., due to competitive effects, unlike in many prior art techniques. Although some drift may occur during the amplification process, the drift may be relatively small. For example, in certain embodiments, the ratios or percentages of a representative unique oligonucleotide sequence, relative to the starting overall population, on the average, may change upon amplification by no more than about 10%, no more than about 8%, no more than about 6%, no more than about 5%, no more than about 4%, no more than about 3%, no more than about 2%, no more than about 1%, etc., relative to the starting ratio or percentage of the oligonucleotide sequence, prior to amplification. However, it should be noted that the oligonucleotide sequence itself, prior to any amplification, may also exhibit some variability, which is not included in the above numbers.

The unique oligonucleotides within a plurality of oligonucleotides may have the same or different lengths. If more than one unique oligonucleotide is present, then the unique oligonucleotides may independently have the same or different lengths. For example, in some cases, a plurality of oligonucleotides may have an average length (number average) of at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 50, at least 60, at least 65, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least

400, or at least 450 nucleotides. In some cases, the average length may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides. Combinations of any of these are also possible, e.g., the average length may be between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc.

In one set of embodiments, any suitable technique may be used to amplify the plurality of oligonucleotides. In some cases, for each oligonucleotide to be amplified, at least about 100, at least about 300, at least about 500, at least about 1,000, at least about 3,000, at least about 5,000, at least about 10,000, at least about 30,000, at least about 50,000 at least about 100,000, at least about 300,000, at least about 500,000, at least about 1,000,000 copies, at least about 3,000,000 copies, at least about 5,000,000 copies, at least about 10,000,000 copies, at least about 30,000,000 copies, at least about 50,000,000 copies, or at least about 100,000,000 copies of the oligonucleotide may be produced using any of the amplification techniques discussed herein (e.g., including PCR amplification, *in vitro* transcription etc.). As discussed, in some cases, the amplification may occur without substantial selective amplification of some oligonucleotide sequences over others.

Any suitable technique may be used to generate the plurality of oligonucleotides. For example, the plurality of oligonucleotides may be synthetically produced, grown within a cell, grown on a substrate (e.g., in an array), or the like. Techniques for producing oligonucleotides are known to those of ordinary skill in the art. The plurality of oligonucleotides may also be computationally designed in some embodiments.

In one embodiment, the oligonucleotides may be amplified using PCR (polymerase chain reaction). In some cases, the oligonucleotides may be amplified while contained in a common liquid or solution. This is to be contrasted with certain PCR techniques, such as emulsion PCR or digital PCR, which requires separation of the oligonucleotides, e.g., into separate compartments or droplets, prior to amplification so as to prevent relatively selective amplification of certain oligonucleotides from

occurring. However, surprisingly, it has been found that such separation is not required, and other techniques (such as is described herein) may be used to prevent or reduce selective amplification while keeping the oligonucleotides together within a common solution.

As mentioned, in some cases, by using certain oligonucleotide structures and/or certain types of PCR techniques, the amount of selective amplification that may occur may be reduced or eliminated. For instance, in one set of embodiments, oligonucleotides are chosen to minimize competitive effects. For instance, the oligonucleotides may have substantially the same lengths, and/or share identical or similar portions or regions.

For example, in one set of embodiments, the oligonucleotides may have a distribution of lengths such that no more than about 10%, no more than about 5%, no more than about 3%, or no more than about 1% of the oligonucleotides has a length that is less than about 80% or greater than about 120%, less than about 90% or greater than about 110%, or less than about 95% or greater than about 105% of the overall average length of the plurality of nucleotides.

In another set of embodiments the oligonucleotides may share one or more regions, such as index regions, that are identical or substantially similar. The oligonucleotides sharing index or other regions may have substantially the same lengths, as discussed above, or different lengths. In some embodiments, the oligonucleotides may comprise at least two index regions that each are identical or substantially similar, surrounding a variable region having different nucleotide sequences, and optionally, different lengths. For example, the oligonucleotides may include, in sequence, a first region that is identical or substantially similar to the other oligonucleotides, a second region that is not identical, and optionally, a third region that is identical or substantially similar to the other oligonucleotides. In some embodiments, competition of oligonucleotides may be controlled by using oligonucleotides selected to reduce amplification bias. For instance, in some cases, groups of oligonucleotides that have similar compositions may be amplified together.

In some cases, the index regions may have a length of greater than 5, 7, 10, 12, 14, 16, 18, or 20 nucleotides, and/or have a length of less than 30, 28, 25, 22, 20, 18, 16, 14, 12, or 10 nucleotides. For instance, the regions that are identical or substantially similar may have a length of between 18 and 22 nucleotides. The regions may be

identical, or differ by no more than 1, 2, 3, 4, or 5 nucleotides (consecutively or non-consecutively) within the region.

In certain embodiments, primer sequences may be added to facilitate the PCR reaction. For example, the primer sequence may include sequences substantially complementary to a region within the oligonucleotides, e.g., an index region. A variety of such sequences suitable for PCR or *in vitro* transcription may be readily obtained commercially.

In some embodiments, the primer sequence may also include other sequences, e.g., promoter sequences or other sequences that may be added to the oligonucleotide during PCR amplification, such as is shown in Fig. 1 with a T7 promoter. Accordingly, oligonucleotides comprising the original sequence and one or more promoter sequences may be produced in certain cases. Besides the T7 promoter, other suitable promoters that may be used include, but are not limited to, T3 promoters or SP6 promoters. Such promoters may be useful, for example, to facilitate transcription to produce RNA, as is discussed in more detail below. In addition, in some embodiments, more than one promoter may be added.

In one set of embodiments, more than one sequence containing a PCR primer may be used, e.g., to amplify different subsets of the plurality of oligonucleotides. If more than one primer-containing sequence is used, the PCR primers contained on each of them may be the same or different. Examples of suitable PCR primers include those described herein. Thus, for example, in one set of embodiments, for example, the plurality of oligonucleotides may include different subpools having different index regions or other regions as discussed above, which may be selectively amplified through the use of an appropriate sequence including a PCR primer. Any suitable number of subpools may be created for a plurality of oligonucleotides. For example, at least 1, 2, 4, 10, 20, 96, 100, or 192 subpools of the plurality oligonucleotides may be selective amplified, through the use of specific PCR primers.

Thus, in some cases, the oligonucleotides may be formed into "pools" or groups or sets of oligonucleotides within the plurality of oligonucleotides that share one or more common features, such as an index region or other identical sequence. For instance, the common feature in a group or set of oligonucleotides may have an identical sequence of nucleotides of at least 5, 7, 10, 12, 14, 16, 18, or 20 nucleotides, and/or less than 30, 28,

25, 22, 20, 18, 16, 14, 12, or 10 nucleotides. For instance, the common region that is identical or substantially similar in a group of oligonucleotides may have a length of between 18 and 22 nucleotides. The common regions may be identical, or differ by no more than 1, 2, 3, 4, or 5 nucleotides (consecutively or non-consecutively) within the region. In some embodiments, each group or pool may contain two (or more) unique index regions that are not used in any other pool, e.g., to reduce the contamination of off-target amplified products from the amplification products of another.

In certain embodiments, PCR amplification may be monitored, and controlled to reduce or minimize selective amplification. For example, in one set of embodiments, real-time PCR techniques may be used. In some embodiments, the extent of the PCR reaction may be monitored or controlled, for example, by illuminating the solution containing the oligonucleotides with suitable light and determining the amount of fluorescence that is present to monitor the PCR reaction. Accordingly, for example, the reaction conditions may be controlled such that the oligonucleotides react in conditions that minimize the amount of selective amplification, for example, by providing an excess of nucleotides, ions (e.g., Mg^{2+}), enzyme, etc. Once the oligonucleotide concentrations have reached the point where competitive effects may start to occur, the reaction may be stopped before significant selective amplification begins.

After amplification as discussed above, the amplified oligonucleotides may then be transcribed to produce RNA. Further amplification may also occur in this step. For instance, in some cases, each oligonucleotide can be used to produce, on the average, at least about 50, at least about 100, at least about 300, at least about 500, at least about 1,000, at least about 3,000, at least about 5,000, at least about 10,000, at least about 30,000, at least about 50,000 at least about 100,000, at least about 300,000, at least about 500,000 or at least about 1,000,000 transcribed RNA molecules. In some cases, the mass of RNA that is produced may be at least about 10, at least about 20, at least about 30, at least about 50, at least about 100, at least about 200, at least about 300, or at least about 500 times the mass of the oligonucleotides. Thus, for example, one microgram of oligonucleotides may be converted into at least 10 micrograms, at least 30 micrograms, or at least 100 micrograms of RNA.

In one set of embodiments, transcription may occur *in vitro* by exposing the amplified oligonucleotides to a suitable RNA polymerase. A variety of RNA

polymerases are available commercially, including T7, T3, or SP6 RNA polymerases. Other non-limiting examples of RNA polymerases include RNA polymerase I, RNA polymerase II, RNA polymerase III, RNA polymerase IV, or RNA polymerase V. The RNA polymerase may arise from any suitable source, e.g., bacteria, viruses, or eukaryotes. In some embodiments, more than one RNA polymerase may be used. In addition, as previously discussed, in some embodiments, the amplified oligonucleotides may include promoter sequences, such as one or more of T7, T3, or SP6 promoter sequences, that can be used to facilitate the transcription process. Those of ordinary skill in the art will be aware of suitable conditions for causing transcription *in vitro* using RNA polymerases.

In some embodiments, the total amplification bias may be reduced by changing the relative amount of amplification produced by the PCR and the *in vitro* transcription. For example, the PCR can be used to produce smaller amounts of DNA than are typically produced in a PCR, to reduce the amplification bias of this process. However, this reduced yield can be compensated in some cases by increasing the duration of the *in vitro* transcription reaction.

The RNA may, in turn, be reverse transcribed to produce DNA. In one set of embodiments, reverse transcription may occur by exposing the RNA to a suitable reverse transcriptase enzyme. In some cases, the reverse transcriptase may be a viral reverse transcriptase, e.g., M-MLV reverse transcriptase, AMV reverse transcriptase, or the like. A variety of reverse transcriptase enzymes are commercially available. Those of ordinary skill in the art will be aware of suitable conditions for causing reverse transcription to occur.

In certain embodiments, reverse transcription may be facilitated through the use of primer-containing sequences, e.g., containing primers for reverse transcription. In some cases, the primer-containing sequences may contain other sequences or entities as well, although this is not necessarily a requirement. The primer-containing sequences may be added at any suitable point, e.g., just before starting transcription reaction. Suitable transcription primers for conducting reverse transcription may be obtained commercially.

In one set of embodiments, the primer-containing sequence may be incorporated into the DNA during production of the DNA by the reverse transcriptase. In some cases,

the primer-containing sequence may contain other entities, and/or sequences suitable for attaching other entities (e.g., on the 5' or 3' ends, internally, etc.). For instance, the primer-containing sequence may contain a non-nucleic acid moiety, such as a digoxigenin moiety, a biotin moiety, etc. located on the 5' end, the 3' end, internally, or the like. In some cases, the signaling entity that can be subsequently detected or determined may be introduced to the DNA. For instance, the signaling entity may be fluorescent, or a specific nucleotide sequence that can be determined, e.g., enzymatically. Examples of signaling entities are discussed in more detail below.

In some cases, the RNA may be purified prior to reverse transcription. However, it should be noted that purification is not required, and in other embodiments, the RNA may be reverse transcribed to form DNA without any intermediate purification steps. If the RNA is purified, it may be purified using any suitable technique, e.g., by passing the RNA over a suitable column to remove oligonucleotides.

Optionally, in some embodiments, the RNA may be separated from the DNA or the DNA may be purified in some fashion. For example, the RNA may be selectively degraded, relative to the DNA. In one set of embodiments, the RNA may be degraded relative to the DNA by alkaline hydrolysis. For instance, the pH of the solution may be raised to at least about 8, at least about 9, at least about 10, etc. Any suitable alkaline may be used to raise the pH. In some cases, after degradation of the RNA, the pH may also be lowered, e.g., to about 7, to about 7.4, to physiological conditions, or the like. In some cases, techniques such as enzymatic degradation can be used to selectively degrade RNA, relative to DNA. The DNA may also be purified using techniques such as column purification, ethanol precipitation, and/or solid-phase reversible immobilization techniques. In addition, in some cases, the DNA may be concentrated, e.g., through evaporation techniques.

In addition, techniques such as those described above may be scaled-up or "numbered-up" to produce larger quantities of material. For example, a process may be repeated using multi-well techniques or by simultaneously running multiple reactions in parallel, etc. to produce larger quantities or masses of oligonucleotides. As a non-limiting example, in one embodiment, processes such as those discussed herein may be performed using multiple wells of a microtiter plate (e.g., having 96, 384, 1536, wells, etc.) to increase output.

The DNA may be used for a variety of purposes, in different embodiments of the invention. For example, in certain embodiments, the DNA may be hybridized to nucleic acid species in liquid samples, e.g., extracted from a variety of biological sources, including human. In some cases, the DNA may be used to physically separate one set of nucleic acids from another, or as primers for PCR or reverse transcription.

In addition, as previously discussed, in certain aspects, signaling entities are incorporated into DNA in some embodiments. The signaling entities may be determined for a variety of purposes. For example, the DNA that is produced may be used as a biological probe, and the signaling entities may be determined in some fashion, e.g., quantitatively or qualitatively, to determine a characteristic or feature of the probe. Examples include, but are not limited to, the position of the probe, the activity of the probe, the concentration of the probe, or the like.

In some cases, signaling entities within a sample may be determined, e.g., spatially, using a variety of techniques. In some embodiments, the signaling entities may be fluorescent, and techniques for determining fluorescence within a sample, such as fluorescence microscopy or confocal microscopy, may be used to spatially identify the positions of signaling entities within a cell. In some cases, the positions of the entities within the sample may be determined in two or even three dimensions.

In some embodiments, the spatial positions of the signaling entities may be determined at relatively high resolutions. For instance, the positions may be determined at spatial resolutions of better than about 100 micrometers, better than about 30 micrometers, better than about 10 micrometers, better than about 3 micrometers, better than about 1 micrometer, better than about 800 nm, better than about 600 nm, better than about 500 nm, better than about 400 nm, better than about 300 nm, better than about 200 nm, better than about 100 nm, better than about 90 nm, better than about 80 nm, better than about 70 nm, better than about 60 nm, better than about 50 nm, better than about 40 nm, better than about 30 nm, better than about 20 nm, or better than about 10 nm, etc.

There are a variety of techniques able to determine or image the spatial positions of entities optically, e.g., using fluorescence microscopy. In some cases, the spatial positions may be determined at super resolutions, or at resolutions better than the wavelength of light. Non-limiting examples include STORM (stochastic optical reconstruction microscopy), STED (stimulated emission depletion microscopy), NSOM

(Near-field Scanning Optical Microscopy), 4Pi microscopy, SIM (Structured Illumination Microscopy), SMI (Spatially Modulated Illumination) microscopy, RESOLFT (Reversible Saturable Optically Linear Fluorescence Transition Microscopy), GSD (Ground State Depletion Microscopy), SSEVI (Saturated Structured-Illumination Microscopy), SPDM (Spectral Precision Distance Microscopy), Photo-Activated Localization Microscopy (PALM), Fluorescence Photoactivation Localization Microscopy (FPALM), LIMON (3D Light Microscopical Nanosizing Microscopy), Super-resolution optical fluctuation imaging (SOFI), or the like. See, e.g., U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.;* U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al;* or Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.,* each incorporated herein by reference in their entireties.

In addition, the signaling entity may be inactivated in some cases. For example, in some embodiments, a first secondary nucleic acid probe containing a signaling entity may be applied to a sample that can recognize a first read sequence, then the first secondary nucleic acid probe can be inactivated before a second secondary nucleic acid probe is applied to the sample. If multiple signaling entities are used, the same or different techniques may be used to inactivate the signaling entities, and some or all of the multiple signaling entities may be inactivated, e.g., sequentially or simultaneously.

Inactivation may be caused by removal of the signaling entity (e.g., from the sample, or from the nucleic acid probe, etc.), and/or by chemically altering the signaling entity in some fashion, e.g., by photobleaching the signaling entity, bleaching or chemically altering the structure of the signaling entity, etc.). For instance, in one set of embodiments, a fluorescent signaling entity may be inactivated by chemical or optical techniques such as oxidation, photobleaching, chemically bleaching, stringent washing or enzymatic digestion or reaction by exposure to an enzyme, dissociating the signaling entity from other components (e.g., a probe), chemical reaction of the signaling entity (e.g., to a reactant able to alter the structure of the signaling entity) or the like.

In some embodiments, various nucleic acid probes (including primary and/or secondary nucleic acid probes) may include one or more signaling entities. If more than

one nucleic acid probe is used, the signaling entities may each by the same or different. In certain embodiments, a signaling entity is any entity able to emit light. For instance, in one embodiment, the signaling entity is fluorescent. In other embodiments, the signaling entity may be phosphorescent, radioactive, absorptive, etc. In some cases, the signaling entity is any entity that can be determined within a sample at relatively high resolutions, e.g., at resolutions better than the wavelength of visible light. The signaling entity may be, for example, a dye, a small molecule, a peptide or protein, or the like. The signaling entity may be a single molecule in some cases. If multiple secondary nucleic acid probes are used, the nucleic acid probes may comprise the same or different signaling entities.

Non-limiting examples of signaling entities include fluorescent entities (fluorophores) or phosphorescent entities, for example, cyanine dyes (e.g., Cy2, Cy3, Cy3B, Cy5, Cy5.5, Cy7, etc.), Alexa Fluor dyes, Atto dyes, photoswtichable dyes, photoactivatable dyes, fluorescent dyes, metal nanoparticles, semiconductor nanoparticles or "quantum dots", fluorescent proteins such as GFP (Green Fluorescent Protein), or photoactivabale fluorescent proteins, such as PAGFP, PSCFP, PSCFP2, Dendra, Dendra2, EosFP, tdEos, mEos2, mEos3, PAmCherry, PAtagRFP, mMaple, mMaple2, and mMaple3. Other suitable signaling entities are known to those of ordinary skill in the art. See, e.g., U.S. Pat. No. 7,838,302 or U.S. Pat. Apl. Ser. No. 61/979,436, each incorporated herein by reference in its entirety.

As used herein, the term "light" generally refers to electromagnetic radiation, having any suitable wavelength (or equivalently, frequency). For instance, in some embodiments, the light may include wavelengths in the optical or visual range (for example, having a wavelength of between about 400 nm and about 700 nm, i.e., "visible light"), infrared wavelengths (for example, having a wavelength of between about 300 micrometers and 700 nm), ultraviolet wavelengths (for example, having a wavelength of between about 400 nm and about 10 nm), or the like. In certain cases, as discussed in detail below, more than one entity may be used, i.e., entities that are chemically different or distinct, for example, structurally. However, in other cases, the entities may be chemically identical or at least substantially chemically identical.

In one set of embodiments, the signaling entity is "switchable," i.e., the entity can be switched between two or more states, at least one of which emits light having a

desired wavelength. In the other state(s), the entity may emit no light, or emit light at a different wavelength. For instance, an entity may be "activated" to a first state able to produce light having a desired wavelength, and "deactivated" to a second state not able to emit light of the same wavelength. An entity is "photoactivatable" if it can be activated by incident light of a suitable wavelength. As a non-limiting example, Cy5, can be switched between a fluorescent and a dark state in a controlled and reversible manner by light of different wavelengths, i.e., 633 nm (or 642nm, 647nm, 656 nm) red light can switch or deactivate Cy5 to a stable dark state, while 405 nm green light can switch or activate the Cy5 back to the fluorescent state. In some cases, the entity can be reversibly switched between the two or more states, e.g., upon exposure to the proper stimuli. For example, a first stimuli (e.g., a first wavelength of light) may be used to activate the switchable entity, while a second stimuli (e.g., a second wavelength of light) may be used to deactivate the switchable entity, for instance, to a non-emitting state. Any suitable method may be used to activate the entity. For example, in one embodiment, incident light of a suitable wavelength may be used to activate the entity to emit light, i.e., the entity is "photoswitchable." Thus, the photoswitchable entity can be switched between different light-emitting or non-emitting states by incident light, e.g., of different wavelengths. The light may be monochromatic (e.g., produced using a laser) or polychromatic. In another embodiment, the entity may be activated upon stimulation by electric field and/or magnetic field. In other embodiments, the entity may be activated upon exposure to a suitable chemical environment, e.g., by adjusting the pH, or inducing a reversible chemical reaction involving the entity, etc. Similarly, any suitable method may be used to deactivate the entity, and the methods of activating and deactivating the entity need not be the same. For instance, the entity may be deactivated upon exposure to incident light of a suitable wavelength, or the entity may be deactivated by waiting a sufficient time.

Typically, a "switchable" entity can be identified by one of ordinary skill in the art by determining conditions under which an entity in a first state can emit light when exposed to an excitation wavelength, switching the entity from the first state to the second state, e.g., upon exposure to light of a switching wavelength, then showing that the entity, while in the second state can no longer emit light (or emits light at a much reduced intensity) when exposed to the excitation wavelength.

In one set of embodiments, as discussed, a switchable entity may be switched upon exposure to light. In some cases, the light used to activate the switchable entity may come from an external source, e.g., a light source such as a laser light source, another light-emitting entity proximate the switchable entity, etc. The second, light emitting entity, in some cases, may be a fluorescent entity, and in certain embodiments, the second, light-emitting entity may itself also be a switchable entity.

In some embodiments, the switchable entity includes a first, light-emitting portion (e.g., a fluorophore), and a second portion that activates or "switches" the first portion. For example, upon exposure to light, the second portion of the switchable entity may activate the first portion, causing the first portion to emit light. Examples of activator portions include, but are not limited to, Alexa Fluor 405 (Invitrogen), Alexa Fluor 488 (Invitrogen), Cy2 (GE Healthcare), Cy3 (GE Healthcare), Cy3B (GE Healthcare), Cy3.5 (GE Healthcare), or other suitable dyes. Examples of light-emitting portions include, but are not limited to, Cy5, Cy5.5 (GE Healthcare), Cy7 (GE Healthcare), Alexa Fluor 647 (Invitrogen), Alexa Fluor 680 (Invitrogen), Alexa Fluor 700 (Invitrogen), Alexa Fluor 750 (Invitrogen), Alexa Fluor 790 (Invitrogen), DiD, DiR, YOYO-3 (Invitrogen), YO-PRO-3 (Invitrogen), TOT-3 (Invitrogen), TO-PRO-3 (Invitrogen) or other suitable dyes. These may linked together, e.g., covalently, for example, directly, or through a linker, e.g., forming compounds such as, but not limited to, Cy5-Alexa Fluor 405, Cy5-Alexa Fluor 488, Cy5-Cy2, Cy5-Cy3, Cy5-Cy3.5, Cy5.5-Alexa Fluor 405, Cy5.5-Alexa Fluor 488, Cy5.5-Cy2, Cy5.5-Cy3, Cy5.5-Cy3.5, Cy7-Alexa Fluor 405, Cy7-Alexa Fluor 488, Cy7-Cy2, Cy7-Cy3, Cy7-Cy3.5, Alexa Fluor 647-Alexa Fluor 405, Alexa Fluor 647-Alexa Fluor 488, Alexa Fluor 647-Cy2, Alexa Fluor 647-Cy3, Alexa Fluor 647-Cy3.5, Alexa Fluor 750-Alexa Fluor 405, Alexa Fluor 750-Alexa Fluor 488, Alexa Fluor 750-Cy2, Alexa Fluor 750-Cy3, or Alexa Fluor 750-Cy3.5. . Those of ordinary skill in the art will be aware of the structures of these and other compounds, many of which are available commercially. The portions may be linked via a covalent bond, or by a linker, such as those described in detail below. Other light-emitting or activator portions may include portions having two quatemized nitrogen atoms joined by a polymethine chain, where each nitrogen is independently part of a heteroaromatic moiety, such as pyrrole, imidazole, thiazole, pyridine, quinoine, indole,

benzothiazole, etc., or part of a nonaromatic amine. In some cases, there may be 5, 6, 7, 8, 9, or more carbon atoms between the two nitrogen atoms.

In certain cases, the light-emitting portion and the activator portions, when isolated from each other, may each be fhiorophores, i.e., entities that can emit light of a certain, emission wavelength when exposed to a stimulus, for example, an excitation wavelength. However, when a switchable entity is formed that comprises the first fluorophore and the second fluorophore, the first fluorophore forms a first, light-emitting portion and the second fluorophore forms an activator portion that switches that activates or "switches" the first portion in response to a stimulus. For example, the switchable entity may comprise a first fluorophore directly bonded to the second fluorophore, or the first and second entity may be connected via a linker or a common entity. Whether a pair of light-emitting portion and activator portion produces a suitable switchable entity can be tested by methods known to those of ordinary skills in the art. For example, light of various wavelength can be used to stimulate the pair and emission light from the light-emitting portion can be measured to determined wither the pair makes a suitable switch.

As a non-limiting example, Cy3 and Cy5 may be linked together to form such an entity. In this example, Cy3 is an activator portion that is able to activate Cy5, the light-emission portion. Thus, light at or near the absorption maximum (e.g., near 532 nm light for Cy3) of the activation or second portion of the entity may cause that portion to activate the first, light-emitting portion, thereby causing the first portion to emit light (e.g., near 647 nm for Cy5). See, e.g., U.S. Pat. No. 7,838,302, incorporated herein by reference in its entirety. In some cases, the first, light-emitting portion can subsequently be deactivated by any suitable technique (e.g., by directing 647 nm red light to the Cy5 portion of the molecule).

Other non-limiting examples of potentially suitable activator portions include 1,5 IAEDANS, 1,8-ANS, 4-Methylumbelliferone, 5-carboxy-2,7-dichlorofluorescein, 5-Carboxyfluorescein (5-FAM), 5-Carboxynapthofluorescein, 5-Carboxytetramethylrhodamine (5-TAMRA), 5-FAM (5-Carboxyfluorescein), 5-HAT (Hydroxy Tryptamine), 5-Hydroxy Tryptamine (HAT), 5-ROX (carboxy-X-rhodamine), 5-TAMRA (5-Carboxytetramethylrhodamine), 6-Carboxyrhodamine 6G, 6-CR 6G, 6-JOE, 7-Amino-4-methylcoumarin, 7-Aminoactinomycin D (7-AAD), 7-Hydroxy-4-methylcoumarin, 9-Amino-6-chloro-2-methoxyacridine, ABQ, Acid Fuchsin, ACMA (9-

Amino-6-chloro-2-methoxyacridine), Acridine Orange, Acridine Red, Acridine Yellow, Acriflavin, Acriflavin Feulgen SITSA, Alexa Fluor 350, Alexa Fluor 405, Alexa Fluor 430, Alexa Fluor 488, Alexa Fluor 500, Alexa Fluor 514, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 555, Alexa Fluor 568, Alexa Fluor 594, Alexa Fluor 610, Alexa Fluor 633, Alexa Fluor 635, Alizarin Complexon, Alizarin Red, AMC, AMCA-S, AMCA (Aminomethylcoumarin), AMCA-X, Aminoactinomycin D, Aminocoumarin, Aminomethylcoumarin (AMCA), Anilin Blue, Anthrocyl stearate, APTRA-BTC, APTS, Astrazon Brilliant Red 4G, Astrazon Orange R, Astrazon Red 6B, Astrazon Yellow 7 GLL, Atabrine, ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 520, ATTO 532, ATTO 550, ATTO 565, ATTO 590, ATTO 594, ATTO 610, ATTO 611X, ATTO 620, ATTO 633, ATTO 635, ATTO 647, ATTO 647N, ATTO 655, ATTO 680, ATTO 700, ATTO 725, ATTO 740, ATTO-TAG CBQCA, ATTO-TAG FQ, Auramine, Aurophosphine G, Aurophosphine, BAO 9 (Bisaminophenyloxadiazole), BCECF (high pH), BCECF (low pH), Berberine Sulphate, Bimane, Bisbenzamide, Bisbenzimide (Hoechst), bis-BTC, Blancophor FFG, Blancophor SV, BOBO -1, BOBO -3, Bodipy 492/515, Bodipy 493/503, Bodipy 500/510, Bodipy 505/515, Bodipy 530/550, Bodipy 542/563, Bodipy 558/568, Bodipy 564/570, Bodipy 576/589, Bodipy 581/591, Bodipy 630/650-X, Bodipy 650/665-X, Bodipy 665/676, Bodipy Fl, Bodipy FL ATP, Bodipy Fl-Ceramide, Bodipy R6G, Bodipy TMR, Bodipy TMR-X conjugate, Bodipy TMR-X, SE, Bodipy TR, Bodipy TR ATP, Bodipy TR-X SE, BO-PRO -1, BO-PRO -3, Brilliant Sulphoflavin FF, BTC, BTC-5N, Calcein, Calcein Blue, Calcium Crimson, Calcium Green, Calcium Green-1 Ca^{2+} Dye, Calcium Green-2 Ca^{2+}, Calcium Green-5N Ca^{2+}, Calcium Green-C18 Ca^{2+}, Calcium Orange, Calcofluor White, Carboxy-X-rhodamine (5-ROX), Cascade Blue, Cascade Yellow, Catecholamine, CCF2 (GeneBlazer), CFDA, Chromomycin A, Chromomycin A, CL-NERF, CMFDA, Coumarin Phalloidin, CPM Methylcoumarin, CTC, CTC Formazan, Cy2, Cy3.1 8, Cy3.5, Cy3, Cy5.1 8, cyclic AMP Fluorosensor (FiCRhR), Dabcyl, Dansyl, Dansyl Amine, Dansyl Cadaverine, Dansyl Chloride, Dansyl DHPE, Dansyl fluoride, DAPI, Dapoxyl, Dapoxyl 2, Dapoxyl 3' DCFDA, DCFH (Dichlorodihydrofluorescein Diacetate), DDAO, DHR (Dihydorhodamine 123), Di-4-ANEPPS, Di-8-ANEPPS (non-ratio), DiA (4-DM6-ASP), Dichlorodihydrofluorescein Diacetate (DCFH), DiD - Lipophilic Tracer, DiD (DiIC18(5)), DIDS, Dihydorhodamine 123 (DHR), DiI (DiIC18(3)), Dinitrophenol, DiO

(DiOC18(3)), DiR, DiR (DiIC18(7)), DM-NERF (high pH), DNP, Dopamine, DTAF, DY-630-NHS, DY-635-NHS, DyLight 405, DyLight 488, DyLight 549, DyLight 633, DyLight 649, DyLight 680, DyLight 800, ELF 97, Eosin, Erythrosin, Erythrosin ITC, Ethidium Bromide, Ethidium homodimer -1 (EthD-1), Euchrysin, EukoLight, Europium (III) chloride, Fast Blue, FDA, Feulgen (Pararosaniline), FIF (Formaldehyd Induced Fluorescence), FITC, Flazo Orange, Fluo-3, Fluo-4, Fluorescein (FITC), Fluorescein Diacetate, Fluoro-Emerald, Fluoro-Gold (Hydroxy stilbamidine), Fluor-Ruby, FluorX, FM 1-43, FM 4-46, Fura Red (high pH), Fura Red/Fluo-3, Fura-2, Fura-2/BCECF, Genacryl Brilliant Red B, Genacryl Brilliant Yellow 10GF, Genacryl Pink 3G, Genacryl Yellow 5GF, GeneBlazer (CCF2), Gloxalic Acid, Granular blue, Haematoporphyrin, Hoechst 33258, Hoechst 33342, Hoechst 34580, HPTS, Hydroxycoumarin, Hydroxystilbamidine (FluoroGold), Hydroxytryptamine, Indo-1, high calcium, Indo-1, low calcium, Indodicarbocyanine (DiD), Indotricarbocyanine (DiR), Intrawhite Cf, JC-1, JO-JO-1, JO-PRO-1, LaserPro, Laurodan, LDS 751 (DNA), LDS 751 (RNA), Leucophor PAF, Leucophor SF, Leucophor WS, Lissamine Rhodamine, Lissamine Rhodamine B, Calcein/Ethidium homodimer, LOLO-1, LO-PRO-1, Lucifer Yellow, Lyso Tracker Blue, Lyso Tracker Blue-White, Lyso Tracker Green, Lyso Tracker Red, Lyso Tracker Yellow, LysoSensor Blue, LysoSensor Green, LysoSensor Yellow/Blue, Mag Green, Magdala Red (Phloxin B), Mag-Fura Red, Mag-Fura-2, Mag-Fura-5, Mag-Indo-1, Magnesium Green, Magnesium Orange, Malachite Green, Marina Blue, Maxilon Brilliant Flavin 10 GFF, Maxilon Brilliant Flavin 8 GFF, Merocyanin, Methoxycoumarin, Mitotracker Green FM, Mitotracker Orange, Mitotracker Red, Mitramycin, Monobromobimane, Monobromobimane (mBBr-GSH), Monochlorobimane, MPS (Methyl Green Pyronine Stilbene), NBD, NBD Amine, Nile Red, Nitrobenzoxadidole, Noradrenaline, Nuclear Fast Red, Nuclear Yellow, Nylosan Brilliant lavin E8G, Oregon Green, Oregon Green 488-X, Oregon Green, Oregon Green 488, Oregon Green 500, Oregon Green 514, Pacific Blue, Pararosaniline (Feulgen), PBFI, Phloxin B (Magdala Red), Phorwite AR, Phorwite BKL, Phorwite Rev, Phorwite RPA, Phosphine 3R, PKH26 (Sigma), PKH67, PMIA, Pontochrome Blue Black, POPO-1, POPO-3, PO-PRO-1, PO-PRO-3, Primuline, Procion Yellow, Propidium Iodid (PI), PyMPO, Pyrene, Pyronine, Pyronine B, Pyrozal Brilliant Flavin 7GF, QSY 7, Quinacrine Mustard, Resorufin, RH 414, Rhod-2, Rhodamine, Rhodamine 110, Rhodamine 123,

Rhodamine 5 GLD, Rhodamine 6G, Rhodamine B, Rhodamine B 200, Rhodamine B extra, Rhodamine BB, Rhodamine BG, Rhodamine Green, Rhodamine Phallicidine, Rhodamine Phalloidine, Rhodamine Red, Rhodamine WT, Rose Bengal, S65A, S65C, S65L, S65T, SBFI, Serotonin, Sevron Brilliant Red 2B, Sevron Brilliant Red 4G, Sevron Brilliant Red B, Sevron Orange, Sevron Yellow L, SITS, SITS (Primuline), SITS (Stilbene Isothiosulphonic Acid), SNAFL calcein, SNAFL-1, SNAFL-2, SNARF calcein, SNARF1, Sodium Green, SpectrumAqua, SpectrumGreen, SpectrumOrange, Spectrum Red, SPQ (6-methoxy-N-(3-sulfopropyl)quinolinium), Stilbene, Sulphorhodamine B can C, Sulphorhodamine Extra, SYTO 11, SYTO 12, SYTO 13, SYTO 14, SYTO 15, SYTO 16, SYTO 17, SYTO 18, SYTO 20, SYTO 21, SYTO 22, SYTO 23, SYTO 24, SYTO 25, SYTO 40, SYTO 41, SYTO 42, SYTO 43, SYTO 44, SYTO 45, SYTO 59, SYTO 60, SYTO 61, SYTO 62, SYTO 63, SYTO 64, SYTO 80, SYTO 81, SYTO 82, SYTO 83, SYTO 84, SYTO 85, SYTOX Blue, SYTOX Green, SYTOX Orange, Tetracycline, Tetramethylrhodamine (TAMRA), Texas Red, Texas Red-X conjugate, Thiadicarbocyanine (DiSC3), Thiazine Red R, Thiazole Orange, Thioflavin 5, Thioflavin S, Thioflavin TCN, Thiolyte, Thiozole Orange, Tinopol CBS (Calcofluor White), TMR, TO-PRO-1, TO-PRO-3, TO-PRO-5, TOTO-1, TOTO-3, TRITC (tetramethylrodamine isothiocyanate), True Blue, TruRed, Ultralite, Uranine B, Uvitex SFC, WW 781, X-Rhodamine, XRITC, Xylene Orange, Y66F, Y66H, Y66W, YO-PRO-1, YO-PRO-3, YOYO-1, YOYO-3, SYBR Green, Thiazole orange (interchelating dyes), or combinations thereof.

In some aspects, the nucleotides can be used to study a sample, such as a biological sample. For instance, the nucleotides may be used to determine nucleic acids within a cell or other sample. The sample may include a cell culture, a suspension of cells, a biological tissue, a biopsy, an organism, or the like. The sample may also be cell-free but nevertheless contain nucleic acids. If the sample contains a cell, the cell may be a human cell, or any other suitable cell, e.g., a mammalian cell, a fish cell, an insect cell, a plant cell, or the like. More than one cell may be present in some cases.

The nucleic acids to be determined may be, for example, DNA, RNA, or other nucleic acids that are present within a cell (or other sample). The nucleic acids may be endogenous to the cell, or added to the cell. For instance, the nucleic acid may be viral, or artificially created. In some cases, the nucleic acid to be determined may be expressed

by the cell. The nucleic acid is RNA in some embodiments. The RNA may be coding and/or non-coding RNA. Non-limiting examples of RNA that may be studied within the cell include mRNA, siRNA, rRNA, miRNA, tRNA, IncRNA, snoRNAs, snRNAs, exRNAs, piRNAs, or the like. In some embodiments, for example, at least some of the plurality of oligonucleotides are complementary to a portion of a specific chromosome sequence, e.g., of a human chromosome.

In some cases, a significant portion of the nucleic acid within the cell may be studied. For instance, in some cases, enough of the RNA present within a cell may be determined so as to produce a partial or complete transcriptome of the cell. In some cases, at least 4 types of mRNAs are determined within a cell, and in some cases, at least 3, at least 4, at least 7, at least 8, at least 12, at least 14, at least 15, at least 16, at least 22, at least 30, at least 31, at least 32, at least 50, at least 63, at least 64, at least 72, at least 75, at least 100, at least 127, at least 128, at least 140, at least 255, at least 256, at least 500, at least 1,000, at least 1,500, at least 2,000, at least 2,500, at least 3,000, at least 4,000, at least 5,000, at least 7,500, at least 10,000, at least 12,000, at least 15,000, at least 20,000, at least 25,000, at least 30,000, at least 40,000, at least 50,000, at least 75,000, or at least 100,000 types of mRNAs may be determined within a cell.

In some cases, the transcriptome of a cell may be determined. It should be understood that the transriptome generally encompasses all RNA molecules produced within a cell, not just mRNA. Thus, for instance, the transcriptome may also include rRNA, tRNA, etc. In some embodiments, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or 100% of the transcriptome of a cell may be determined.

The determination of one or more nucleic acids within the cell or other sample may be qualitative and/or quantitative. In addition, the determination may also be spatial, e.g., the position of the nucleic acid within the cell or other sample may be determined in two or three dimensions. In some embodiments, the positions, number, and/or concentrations of nucleic acids within the cell (or other sample) may be determined.

One non-limiting example of such as system may be found in U.S. Provisional Patent Application Serial No. 62/031,062, filed July 30, 2014, entitled "Systems and

Methods for Determining Nucleic Acids," by Zhuang, *et al.* incorporated herein by reference in its entirety.

The following documents are each incorporated herein by reference in their entireties: U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.;* U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.;* and Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.* In addition, incorporated herein by reference in their entireties are U.S. Provisional Patent Application Serial No. 62/031,062, filed July 30, 2014, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.;* U.S. Provisional Patent Application Serial No. 62/050,636, filed September 15, 2014, entitled "Probe Library Construction," by Zhuang, *et al.;* U.S. Provisional Patent Application Serial No. 62/142,653, filed April 3, 2015, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al. ;* and a PCT application filed on even date herewith, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*

The following examples are intended to illustrate certain embodiments of the present invention, but do not exemplify the full scope of the invention.

EXAMPLE 1

This example illustrates high-throughput hybridization construction of DNA probes, according to certain embodiments of the invention.

Overview. This protocol uses complex libraries of oligonucleotides as templates for the enzymatic construction of large quantities of single-stranded DNA molecules that can be chemically labeled and which are designed to be hybridized to specific sets of nucleic acid targets that can vary significantly in complexity, i.e. the number of unique target sequences. The use of this protocol involves several basic steps: 1) computational design and optimization of a set of oligonucleotide sequences that will serve as the hybridization regions on the sample of interest; 2) computational design and optimization of a large set of short oligonucleotide sequences to serve as highly specific PCR primers; 3) computational construction of the template molecules; 4) synthesis of the template molecules to create the template library; 5) selection of a sub-set of *in vitro* template

molecules from a template library via PCR; 6) m-viiro-transcription-based amplification of the *in vitro* template molecules into RNA, which serves as the final template for the hybridization probes; 7) reverse transcription of the RNA back into DNA using chemically-modified primers; 8) removal of the RNA via alkaline hydrolysis; and 9) purification of the chemically-modified ssDNA molecules, called probes in this example. These steps are discussed in detail below.

Construction of Probes. Selection of Hybridization Regions. The software OligoArray 2.0 was used to design a large set of potential hybridization regions in the *E. coli* transcriptome in this example. Briefly, this software selects hybridization regions that lie within user-specified ranges for length and melting temperature. This software also screens for off-target hybridization as well as potential secondary structure. This software was used to generate hybridization regions of 30-nt length, a melting-temperature range of 80-85 °C, a GC content of between 50-60%, and a secondary structure melting threshold and cross-hybridization-melting temperature of 75 °C using all annotated, transcribed mRNAs in *E. coli* (K-12 mgl655; NC90013.2).

Design of Hundreds of Orthogonal Primers. The general hybridization probe set may require far fewer unique sequences than is provided in the complex sets of oligonucleotides typically generated by array-based synthesis—the *oligopool.* To exploit this disparity in complexity and significantly lower the cost per experiment, this example uses a method to embed a large number of unique template sets within a single oligopool. Briefly, each template molecule was flanked by a unique pair of PCR primers common only to the template molecules for all probes within a given set. To facilitate the embedding of hundreds of unique probe template sets within a single oligopool, a protocol was created for constructing hundreds of orthogonal PCR primers.

Specifically, the protocol starts by truncating the members of an existing library of 240,000, semi-orthogonal, 25-mer oligonucleotides to 20-nt length and selecting oligos on parameters optimal for PCR: a narrow range of predicted melting-temperature (65-70 °C) and GC content (50-60%); the absence of contiguous runs of the same base longer than 4, i.e. AAAA; and the presence of a 3' GC-clamp, i.e. 2-3 G/C within the final 5 nt. BLAST was then run with these optimized primers against all annotated RNAs in the *E. coli* transcriptome as well as the T7 promoter (TAATACGACTCACTATAGGG) (SEQ ID NO. 1) and a common priming region (P9:

CAGGCATCCGAGAGGTCTGG) (SEQ ID NO. 2) and potential primers with hits to the transcriptome with 12-nt or longer of homology, or with a hit of any length within 3-nt of the 3' end of the potential primer, were removed. Finally, the remaining primers were screened for homology to each other, again using BLAST. Any primer with longer than 11-nt of homology or with any homology within 3-nt of the 3' end of another primer was removed. These cuts reduced the original 240,000 oligos to 198 highly optimized primers for the *E. coli* transcriptome. The final set of 198 primers are listed in Fig. 3.

If required, more primers could be generated using established techniques to create a larger set of initial oligonucleotides, or by relaxing the stringency of the cuts described above. Finally, by changing the transcriptome used to screen the primers, this approach can also be generalized to the generation of optimal index primers for any organism.

Template Construction. To design the template libraries used to create our probe sets, the desired RNA targets were first selected to be stained simultaneously. Individual template molecules were designed by concatenating the following sequences: i) the first of two unique primers for the appropriate mRNA group, ii) the common primer P9, iii) the site for the nicking enzyme Nb.BsmI, iv) the reverse complement of the hybridization region to the target, v) the reverse complement of the nicking enzyme Nb.BsrDI, and vi) the reverse complement of the second unique primer for the given mRNA group. Fig. 2 demonstrates this organization. Multiple probe template sets were combined into large oligopools and these pools were synthesized via CustomArray.

Fig. 2 shows an example template sequence containing a probe the mRNA, acnB. Underlined at the beginning is the sequence of the first primer, not underlined is the common P9 priming site, then underlined is the Nb.BsmI site, not underlined is the reverse complement to the hybridization region for acnB, next underlined is the reverse complement of the Nb.BsrDI site, and the final not underlined portion is the second unique primer. This template is one of 736 used to create probes to stain all mRNAs expressed in the copy number range of 1-10 per cell transcribed from the *E. coli* genomic locus corresponding to base pairs 1-100 kb. See Fig. 4 for the sequences of the 736 probes.

Index PCR. The template for specific probe sets were selected from the complex oligopool via limited-cycle PCR. 0.5 to 1 ng of the complex oligopool was combined with 0.5 micromolar of each primer. The forward primer matched the priming sequence for the desired sub set while the reverse primer was a 5'concatenation of this sequence with the T7 promoter. To avoid the generation of G-quadruplets, which can be difficult to synthesize, the terminal Gs required in the T7 promoter were generated from Gs located at the 5' of the priming region where appropriate. All primers were synthesized by IDT. A 50 microliter reaction volume was amplified either using the KAPA real-time library amplification kit (KAPA Biosystems; KK2701) or via a homemade qPCR mix which included 0.8X EvaGreen (Biotum; 31000-T) and the hot-start Phusion polymerase (New England Biolabs; M0535S). Amplification was followed in real time using Agilent's MX300P or Biorad's CFX Connect. Individual samples were removed before the plateau in amplification, often at concentrations about 10-fold lower than would correspond to this plateau, to minimize distortion of template abundance due to over-amplification. Individual templates were purified with columns according to the manufacturer's instructions (Zymo DNA Clean and Concentrator; D4003) and eluted in RNase-free deionized water.

Amplification via *in vitro* transcription. The template was then amplified via *in vitro* transcription. Briefly, 0.5 to 1 micrograms of template DNA was amplified into 100 to 200 micrograms of RNA in a single 20-30 microliter reaction with a high yield RNA polymerase (New England Biolabs; E2040S). Reactions were supplemented with IX RNase inhibitor (Promega RNasin; N261 1). Amplification was typically run for 2 to 4 hours at 37 °C to maximize the yield. The RNA was not purified after the reaction and was either stored at -80 °C or immediately converted into DNA as described below.

Reverse Transcription. 1-2 nmol of fluorescently-labeled ssDNA probe was created from the above *in vitro* transcription reactions using the reverse transcriptase Maxima H- (Thermo Scientific; EP0751). This enzyme was used because of its higher processivity and temperature resistance, which allowed the conversion of large quantities of RNA into DNA within small volumes at temperatures that disfavor secondary structure formation. The unpurified RNA created above was supplemented with 1.6 mM of each dNPT, 1-2 nmol of fluorescently labeled P9 primer, 300 units of Maxima H-, 60

units of RNasin, and a final IX concentration of the Maxima RT buffer. The final 75 microliter volume was incubated at 50 °C for 60 minutes.

Strand Selection and Purification. The template RNA in the reaction above was then removed from the DNA via alkaline hydrolysis. 75 microliters of a solution of 0.25 M EDTA and 0.5 N NaOH was added to each reverse transcription reaction, and the sample was incubated at 95 °C for 10 minutes. The reaction was immediately neutralized by purifying the ssDNA probe with a modified version of the Zymo Oligo Clean and Concentrator protocol. Specifically, the 5-microgram capacity column was replaced with a 100- microgram capacity DNA column as appropriate. The remainder of the protocol was run according to the manufacturer's instructions. Probe was eluted in 100 microliter RNase-free deionized water and evaporated in a vacuum concentrator. The final pellet was resuspend in 10 microliter RNase-free water and stored at -20 °C. Denaturing poly-acrylimid gel electrophoresis and absorption spectroscopy revealed that this protocol typically produces 90-100% incorporate of the fluorescent primer into full length probe and 60-75% recovery of the total fluorescent probe. Thus, without exceeding a 150-microliter reaction volume, this protocol can be used to create ~2 nmol of fluorescent probe. The small reaction volumes were conducive to the use of high-throughput fluid handling techniques and significantly lower the cost of this reaction as compared to alternative approaches. Thus, 24-96 probes could be constructed in parallel, with minimal hands-on time, across two days, for a final cost of ~$14 per 2 nmol of each probe set.

While several embodiments of the present invention have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the functions and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the present invention. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the teachings of the present invention is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine

experimentation, many equivalents to the specific embodiments of the invention described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, the invention may be practiced otherwise than as specifically described and claimed. The present invention is directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles, materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the scope of the present invention.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

The indefinite articles "a" and "an," as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean "at least one."

The phrase "and/or," as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with "and/or" should be construed in the same fashion, i.e., "one or more" of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the "and/or" clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to "A and/or B", when used in conjunction with open-ended language such as "comprising" can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, "or" should be understood to have the same meaning as "and/or" as defined above. For example, when separating items in a list, "or" or "and/or" shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such

as "only one of or "exactly one of," or, when used in the claims, "consisting of," will refer to the inclusion of exactly one element of a number or list of elements. In general, the term "or" as used herein shall only be interpreted as indicating exclusive alternatives (i.e. "one or the other but not both") when preceded by terms of exclusivity, such as "either," "one of," "only one of," or "exactly one of." "Consisting essentially of," when used in the claims, shall have its ordinary meaning as used in the field of patent law.

As used herein in the specification and in the claims, the phrase "at least one," in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase "at least one" refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, "at least one of A and B" (or, equivalently, "at least one of A or B," or, equivalently "at least one of A and/or B") can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

In the claims, as well as in the specification above, all transitional phrases such as "comprising," "including," "carrying," "having," "containing," "involving," "holding," "composed of," and the like are to be understood to be open-ended, i.e., to mean including but not limited to. Only the transitional phrases "consisting of and "consisting essentially of shall be closed or semi-closed transitional phrases,

respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

CLAIMS

1. A method, comprising:

amplifying at least some of a plurality of oligonucleotides using real-time PCR to produce amplified oligonucleotides;

transcribing *in vitro* at least some of the amplified oligonucleotides to produce RNA;

reverse transcribing the RNA to produce transcribed DNA; and

selectively degrading the RNA relative to the transcribed DNA.

2. The method of claim 1, wherein the plurality of oligonucleotides have an average length of between 10 and 200 nucleotides.

3. The method of any one of claims 1 or 2, wherein the plurality of oligonucleotides includes at least 10 unique oligonucleotide sequences.

4. The method of any one of claims 1-3, wherein the plurality of oligonucleotides includes at least 100 unique oligonucleotide sequences.

5. The method of any one of claims 1-4, wherein the plurality of oligonucleotides includes at least 1,000 unique oligonucleotide sequences.

6. The method of any one of claims 1-5, wherein the plurality of oligonucleotides includes at least 10,000 unique oligonucleotide sequences.

7. The method of any one of claims 1-6, wherein the plurality of oligonucleotides includes at least 100,000 unique oligonucleotide sequences.

8. The method of any one of claims 1-7, wherein the unique oligonucleotide sequences each comprise a first region that is identical in the unique oligonucleotide sequences, and a second region that is not identical in the unique oligonucleotide sequences.

9. The method of claim 8, wherein the unique oligonucleotide sequences each comprise a third region that is identical in the unique oligonucleotide sequences, wherein the second region separates the first region and the third region.

10. The method of any one of claims 1-9, wherein amplifying at least some of the plurality of oligonucleotides comprises exposing a least some of the plurality of oligonucleotides to primer-containing sequences.

11. The method of claim 10, wherein at least some of the primer-containing sequences further comprises a promoter, whereby the promoter is added to at least some of the amplified oligonucleotides.

12. The method of any one of claims 1-10, wherein at least some of the amplified oligonucleotides contain a promoter.

13. The method of any one of claims 11 or 12, wherein the promoter is a T7 promoter.

14. The method of any one of claims 11 or 12, wherein the promoter is a T3 promoter.

15. The method of any one of claims 11 or 12, wherein the promoter is a SP6 promoter.

16. The method of any one of claims 1-15, wherein at least some of the plurality of oligonucleotides comprise at least a first set of oligonucleotides having a first common index region, and a second set of oligonucleotides having a second common index region distinguishable from the first common index region.

17. The method of claim 16, comprising amplifying the first set of oligonucleotides but not the second set of oligonucleotides.

18. The method of any one of claims 16 or 17, wherein the plurality of oligonucleotides comprises at least 2 sets of oligonucleotides having distinguishable common index regions.

19. The method of any one of claims 16-18, wherein the plurality of oligonucleotides comprises at least 4 sets of oligonucleotides having distinguishable common index regions.

20. The method of any one of claims 16-19, wherein the plurality of oligonucleotides comprises at least 10 sets of oligonucleotides having distinguishable common index regions.

21. The method of any one of claims 16-20, wherein the plurality of oligonucleotides comprises at least 20 sets of oligonucleotides having distinguishable common index regions.

22. The method of any one of claims 16-21, wherein the plurality of oligonucleotides comprises at least 96 sets of oligonucleotides having distinguishable common index regions.

23. The method of any one of claims 16-22, wherein the plurality of oligonucleotides comprises at least 100 sets of oligonucleotides having distinguishable common index regions.

24. The method of any one of claims 16-23, wherein the plurality of oligonucleotides comprises at least 192 sets of oligonucleotides having distinguishable common index regions.

25. The method of any one of claims 1-24, wherein amplifying at least some of the plurality of oligonucleotides comprises producing amplified oligonucleotides

comprising one oligonucleotide of the plurality of oligonucleotides and a promoter.

26. The method of any one of claims 1-27, wherein transcribing at least some of the amplified oligonucleotides comprises a mass of RNA that is at least 100-fold greater than the mass of amplified oligonucleotides.

27. The method of any one of claims 1-26, wherein transcribing at least some of the amplified oligonucleotides comprises exposing the amplified oligonucleotides to an RNA polymerase.

28. The method of claim 27, wherein the RNA polymerase is a T7 RNA polymerase.

29. The method of claim 27, wherein the RNA polymerase is a T3 RNA polymerase.

30. The method of claim 27, wherein the RNA polymerase is a SP6 RNA polymerase.

31. The method of claim 27, wherein the RNA polymerase is RNA polymerase I, RNA polymerase II, RNA polymerase III, RNA polymerase IV, or RNA polymerase V.

32. The method of claim 27, wherein the RNA polymerase is a viral RNA polymerase.

33. The method of any one of claims 1-32, wherein transcribing at least some of amplified oligonucleotides to produce RNA comprises producing, on average, at least 10 RNA copies of each of the amplified oligonucleotides.

34. The method of any one of claims 1-33, wherein transcribing at least some of amplified oligonucleotides to produce RNA comprises producing, on average, at least 100 RNA copies of each of the amplified oligonucleotides.

35. The method of any one of claims 1-34, wherein transcribing at least some of amplified oligonucleotides to produce RNA comprises producing, on average, at least 1,000 RNA copies of each of the amplified oligonucleotides.

36. The method of any one of claims 1-35, wherein transcribing at least some of amplified oligonucleotides to produce RNA comprises producing, on average, at least 10,000 RNA copies of each of the amplified oligonucleotides.

37. The method of any one of claims 1-36, wherein reverse transcribing the RNA comprises exposing the RNA to a reverse transcriptase.

38. The method of claim 37, wherein the reverse transcriptase is a viral reverse transcriptase.

39. The method of any one of claims 1-38, wherein reverse transcribing the RNA to produce transcribed DNA occurs without first purifying the RNA from components used to produce the RNA.

40. The method of any one of claims 1-39, further comprising purifying the RNA from components used to produce the RNA prior to reverse transcribing the RNA to produce transcribed DNA.

41. The method of any one of claims 1-40, wherein reverse transcribing the RNA to produce transcribed DNA comprises reverse transcribing the RNA to produce transcribed DNA using a sequence containing a transcription primer.

42. The method of claim 41, wherein the sequence containing a transcription primer is incorporated into the transcribed DNA.

43. The method of any one of claims 41 or 42, wherein the sequence containing a transcription primer further comprises a biotin moiety.

44. The method of any one of claims 41-43, wherein the sequence containing a transcription primer further comprises a digoxigenin moiety.

45. The method of any one of claims 41-44, wherein the sequence containing a transcription primer further comprises a fluorophore.

46. The method of any one of claims 41-45, wherein the sequence containing a transcription primer further comprises a non-nucleic acid moiety.

47. The method of any one of claims 41-46, wherein the sequence containing a transcription primer further comprises a signaling entity attached thereto.

48. The method of claim 47, wherein the signaling entity is fluorescent.

49. The method of any one of claims 47 or 48, wherein the signaling entity is photoswitchable between a first state able to emit light at a first wavelength, and a second state not able to emit light at the first wavelength.

50. The method of any one of claims 47-49, wherein the signaling entity is photactivatable.

51. The method of claim 47, wherein the signaling entity is a nucleic acid sequence.

52. The method of claim 51, wherein the additional nucleic acid sequence is able to hybridize selectively to a second nucleic acid sequence containing an additional signaling entity via Watson-Crick base-pairing.

53. The method of any one of claims 1-52, wherein selectively degrading the RNA relative to the transcribed DNA comprises chemically reducing the RNA.

54. The method of any one of claims 1-53, wherein selectively degrading the RNA relative to the transcribed DNA comprises exposing the RNA to a pH of at least about 9.

55. The method of any one of claims 1-54, wherein selectively degrading the RNA relative to the transcribed DNA comprises subjecting the RNA to alkaline hydrolysis.

56. The method of any one of claims 1-55, wherein selectively degrading the RNA relative to the transcribed DNA comprises enzymatically digesting the RNA.

57. The method of any one of claims 1-56, wherein the transcribed DNA is substantially single-stranded.

58. The method of any one of claims 1-57, wherein the steps are performed in the order recited.

59. The method of any one of claims 1-58, further comprising purifying the transcribed DNA.

60. The method of claim 59, comprising purifying the transcribed DNA by column purification.

61. The method of any one of claims 59 or 60, comprising purifying the transcribed DNA by ethanol precipitation.

62. The method of any one of claims 59-61, comprising purifying the transcribed DNA by solid-phase reversible immobilization.

63. The method of any one of claims 1-62, further comprising concentrating the transcribed DNA.

64. The method of claim 63, comprising concentrating the transcribed DNA via evaporation.

65. The method of any one of claims 1-64, wherein the plurality of oligonucleotides is computationally designed.

66. The method of any one of claims 1-65, wherein the plurality of oligonucleotides is synthetically produced.

67. The method of any one of claims 1-66, wherein the plurality of oligonucleotides arises from a cell.

68. The method of any one of claims 1-67, wherein the plurality of oligonucleotides is produced on a substrate.

69. The method of claim 68, wherein the plurality of oligonucleotides is produced in an array on a substrate.

70. The method of any one of claims 1-69, further comprising separating a subset of the DNA.

71. The method of any one of claims 1-70, wherein each oligonucleotide of a subset of the oligonucleotides comprises an index portion that is identical.

72. The method of any one of claims 1-71, wherein each transcribed DNA of a subset of the oligonucelotides comprises an index portion that is identical, the method further comprising separating the subset of the transcribed DNA based on the identical index portions.

73. A method, comprising:

simultaneously amplifying at least some of a plurality of oligonucleotides in a common solution using PCR to produce amplified oligonucleotides;

transcribing *in vitro* at least some of the amplified oligonucleotides to produce RNA;

reverse transcribing the RNA to produce transcribed DNA; and

selectively degrading the RNA relative to the transcribed DNA.

74. The method of claim 73, wherein the plurality of oligonucleotides have an average length of between 10 and 200 nucleotides.

75. The method of any one of claims 73 or 74, wherein the plurality of oligonucleotides includes at least 10 unique oligonucleotide sequences.

76. The method of any one of claims 73-75, wherein the plurality of oligonucleotides includes at least 100 unique oligonucleotide sequences.

77. The method of any one of claims 73-76, wherein the plurality of oligonucleotides includes at least 1,000 unique oligonucleotide sequences.

78. The method of any one of claims 73-77, wherein the plurality of oligonucleotides includes at least 10,000 unique oligonucleotide sequences.

79. The method of any one of claims 73-78, wherein the plurality of oligonucleotides includes at least 100,000 unique oligonucleotide sequences.

80. The method of any one of claims 73-79, wherein amplifying at least some of the plurality of oligonucleotides comprises exposing at least some of the plurality of oligonucleotides to primer-containing sequences.

81. The method of claim 80, wherein at least some of the primer-containing sequences further comprises a promoter, whereby the promoter is added to at least some of the amplified oligonucleotides.

82. The method of any one of claims 1-80, wherein at least some of the amplified oligonucleotides contain a promoter.

83. The method of any one of claims 81or 82, wherein the promoter is a T7 promoter.

84. The method of any one of claims 81 or 82, wherein the promoter is a T3 promoter.

85. The method of any one of claims 81or 82, wherein the promoter is a SP6 promoter.

86. The method of any one of claims 73-85, wherein at least some of the plurality of oligonucleotides comprise at least a first set of oligonucleotides having a first common index region, and a second set of oligonucleotides having a second common index region distinguishable from the first common index region.

87. The method of claim 86, comprising amplifying the first set of oligonucleotides but not the second set of oligonucleotides.

88. The method of any one of claims 86 or 87, wherein the plurality of oligonucleotides comprises at least 2 sets of oligonucleotides having distinguishable common index regions.

89. The method of any one of claims 86-88, wherein the plurality of oligonucleotides comprises at least 4 sets of oligonucleotides having distinguishable common index regions.

90. The method of any one of claims 86-89, wherein the plurality of oligonucleotides comprises at least 10 sets of oligonucleotides having distinguishable common index regions.

91. The method of any one of claims 86-90, wherein the plurality of oligonucleotides comprises at least 20 sets of oligonucleotides having distinguishable common index regions.

92. The method of any one of claims 86-91, wherein the plurality of oligonucleotides comprises at least 96 sets of oligonucleotides having distinguishable common index regions.

93. The method of any one of claims 86-92, wherein the plurality of oligonucleotides comprises at least 100 sets of oligonucleotides having distinguishable common index regions.

94. The method of any one of claims 86-93, wherein the plurality of oligonucleotides comprises at least 192 sets of oligonucleotides having distinguishable common index regions.

95. The method of any one of claims 73-94, wherein amplifying at least some of the plurality of oligonucleotides comprises producing amplified oligonucleotides comprising one oligonucleotide of the plurality of oligonucleotides and a promoter.

96. The method of any one of claims 73-95, wherein transcribing at least some of the amplified oligonucleotides comprises a mass of RNA that is at least 100-fold greater than the mass of amplified oligonucleotides.

97. The method of any one of claims 73-96, wherein transcribing at least some of the amplified oligonucleotides comprises exposing the amplified oligonucleotides to an RNA polymerase.

98. The method of claim 97, wherein the RNA polymerase is a T7 RNA polymerase.

99. The method of claim 97, wherein the RNA polymerase is a T3 RNA polymerase.

100. The method of claim 97, wherein the RNA polymerase is a SP6 RNA polymerase.

101. The method of claim 97, wherein the RNA polymerase is RNA polymerase I, RNA polymerase II, RNA polymerase III, RNA polymerase IV, or RNA polymerase V.

102. The method of claim 97, wherein the RNA polymerase is a viral RNA polymerase.

103. The method of any one of claims 73-102, wherein transcribing at least some of amplified oligonucleotides to produce RNA comprises producing, on average, at least 10 RNA copies of each of the amplified oligonucleotides.

104. The method of any one of claims 73-103, wherein transcribing at least some of amplified oligonucleotides to produce RNA comprises producing, on average, at least 100 RNA copies of each of the amplified oligonucleotides.

105. The method of any one of claims 73-104, wherein transcribing at least some of amplified oligonucleotides to produce RNA comprises producing, on average, at least 1,000 RNA copies of each of the amplified oligonucleotides.

106. The method of any one of claims 73-105, wherein transcribing at least some of amplified oligonucleotides to produce RNA comprises producing, on average, at least 10,000 RNA copies of each of the amplified oligonucleotides.

107. The method of any one of claims 73-106, wherein reverse transcribing the RNA comprises exposing the RNA to a reverse transcriptase.

108. The method of claim 107, wherein the reverse transcriptase is a viral reverse transcriptase.

109. The method of any one of claims 73-108, wherein reverse transcribing the RNA to produce transcribed DNA occurs without first purifying the RNA from components used to produce the RNA.

110. The method of any one of claims 73-109, further comprising purifying the RNA from components used to produce the RNA prior to reverse transcribing the RNA to produce transcribed DNA.

111. The method of any one of claims 73-110, wherein reverse transcribing the RNA to produce transcribed DNA comprises reverse transcribing the RNA to produce transcribed DNA using a sequence containing a transcription primer.

112. The method of claim 111, wherein the sequence containing a transcription primer is incorporated into the transcribed DNA.

113. The method of any one of claims 111 or 112, wherein the sequence containing a transcription primer further comprises a biotin moiety.

114. The method of any one of claims 111-113, wherein the sequence containing a transcription primer further comprises a digoxigenin moiety.

115. The method of any one of claims 111-114, wherein the sequence containing a transcription primer further comprises a fluorophore.

116. The method of any one of claims 111-115, wherein the sequence containing a transcription primer further comprises a non-nucleic acid moiety.

117. The method of any one of claims 111-116, wherein the sequence containing a transcription primer further comprises a signaling entity attached thereto.

118. The method of claim 117, wherein the signaling entity is fluorescent.

119. The method of any one of claims 117 or 118, wherein the signaling entity is photo switchable between a first state able to emit light at a first wavelength, and a second state not able to emit light at the first wavelength.

120. The method of any one of claims 117-1 19, wherein the signaling entity is photactivatable.

121. The method of claim 120, wherein the signaling entity is a nucleic acid sequence.

122. The method of claim 121, wherein the additional nucleic acid sequence is able to hybridize selectively to a second nucleic acid sequence containing an additional signaling entity via Watson-Crick base-pairing.

123. The method of any one of claims 73-122, wherein selectively degrading the RNA relative to the transcribed DNA comprises chemically reducing the RNA.

124. The method of any one of claims 73-123, wherein selectively degrading the RNA relative to the transcribed DNA comprises exposing the RNA to a pH of at least about 9.

125. The method of any one of claims 73-124, wherein selectively degrading the RNA relative to the transcribed DNA comprises subjecting the RNA to alkaline hydrolysis.

126. The method of any one of claims 73-125, wherein selectively degrading the RNA relative to the transcribed DNA comprises enzymatically digesting the RNA.

127. The method of any one of claims 73-126, wherein the transcribed DNA is substantially single-stranded.

128. The method of any one of claims 73-127, wherein the steps are performed in the order recited.

129. The method of any one of claims 73-128, further comprising purifying the transcribed DNA.

130. The method of claim 129, comprising purifying the transcribed DNA by column purification.

131. The method of any one of claims 129 or 130, comprising purifying the transcribed DNA by ethanol precipitation.

132. The method of any one of claims 129-131, comprising purifying the transcribed DNA by solid-phase reversible immobilization.

133. The method of any one of claims 73-132, further comprising concentrating the transcribed DNA.

134. The method of claim 133, comprising concentrating the transcribed DNA via evaporation.

135. The method of any one of claims 73-134, wherein the plurality of oligonucleotides is computationally designed.

136. The method of any one of claims 73-135, wherein the plurality of oligonucleotides is synthetically produced.

137. The method of any one of claims 73-136, wherein the plurality of oligonucleotides arises from a cell.

138. The method of any one of claims 73-137, wherein the plurality of oligonucleotides is produced on a substrate.

139. The method of claim 138, wherein the plurality of oligonucleotides is produced in an array on a substrate.

140. The method of any one of claims 73-139, further comprising separating a subset of the DNA.

141. The method of any one of claims 73-140, wherein each oligonucleotide of a subset of the oligonucleotides comprises an index portion that is identical.

142. The method of any one of claims 73-141, wherein each transcribed DNA of a subset of the oligonucelotides comprises an index portion that is identical, the method further comprising separating the subset of the transcribed DNA based on the identical index portions.

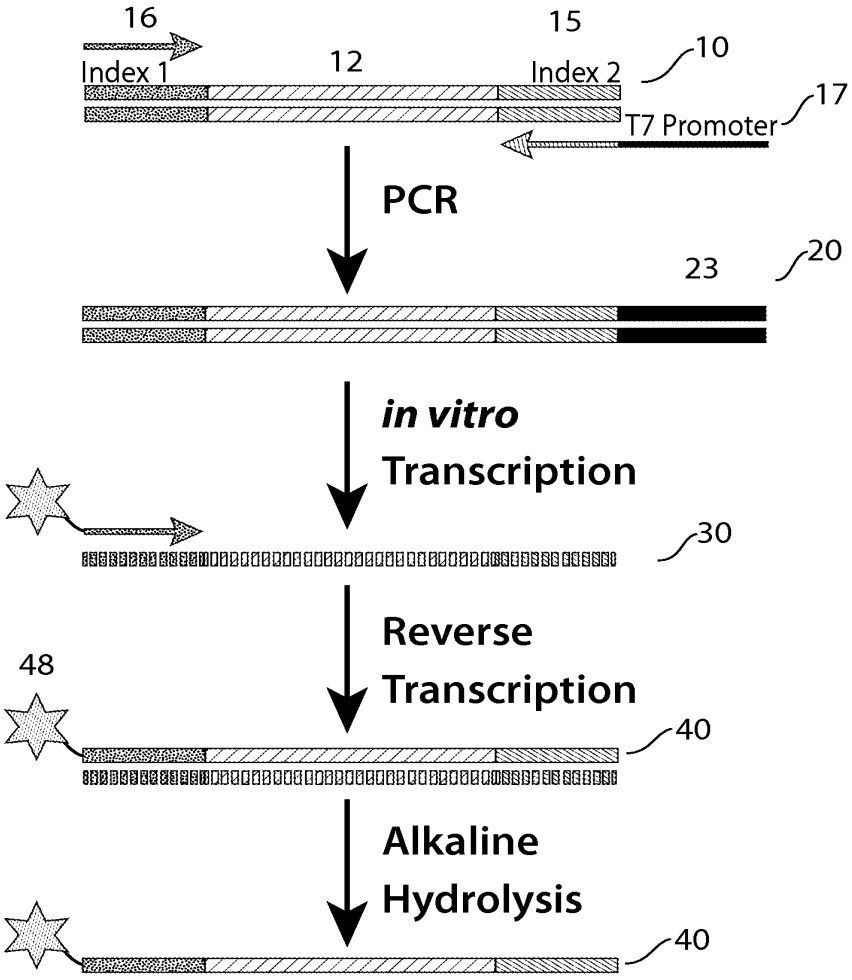


FIG. 1

ACAACCGCGTGTTACAAGGCCAGGCATCCGAGAGGTCTGGGAATGCC AGCAGCTCTACAAGTGCGGCCATTTGGTTCATTGCGCTTCTCCGGAAAC ATTAGC (SEQ ID NO: 3)

FIG. 2

>bc25mer_8982-20
CGGTAGGACCCTACACATCC

>bc25mer_37497-20
TGTGGCGCCTAACCATACAC

>bc25mer_48272-20
GACACCGAGCTAGAATTCGG

>bc25mer_96434-20
TGTCTTGGGTCCTCTTACGG

>bc25mer_135960-20
TATCACGTGAGGCTCCGTGT

>bc25mer_217459-20
GTCTACGGATCCTTGCGTGT

>bc25mer_5177-20
ACGGGTACATTGGATCCACT

>bc25mer_7835-20
AAGACGTGGTCCACCCTGAT

>bc25mer_13747-20
TGCCAGGTCCACTACTTGGG

>bc25mer_14156-20
GACCCGAAGGAGAATGCAAA

>bc25mer_14649-20
ATGTCGTCTCTCAGGGCTTT

>bc25mer_16849-20
AAGGGCGGAAGCACTCTCTT

>bc25mer_25227-20
TGGTCCCGCATTTGTACCTT

>bc25mer_45585-20
AAGTCCGTCTGCTAAGCCAT

>bc25mer_46295-20
CTTCCAGAGTCCTGGCATGA

>bc25mer_46396-20
TCTAACGAGCTCCGTGGTTG

>bc25mer_55377-20
CCTGGGCAGGAAACAACACT

>bc25mer_57092-20
GCGCACACCTTAGTCATAGC

>bc25mer_69827-20
GTGCGATTGTCACATGTGCT

>bc25mer_85616-20
GGCTTGTCCTCAATTATCCG

>bc25mer_90788-20
CCCACGTGACTTGACTTCTC

>bc25mer_108211-20
CCTCTACATGCTCCGGATCC

>bc25mer_113656-20
GAATAGTAACCCGGTCGCAA

>bc25mer_121326-20
TGGGCCACTAGTCTTCGCTA

>bc25mer_123881-20
CGGCACGTACCCTTTACAGG

>bc25mer_138428-20
GATCAGTGAGCCTCACCAAG

>bc25mer_140203-20
TGGTATGTGAGCTCTTTGGG

>bc25mer_159878-20
AGAGTCACGCATACTTGGGA

>bc25mer_163235-20
GGGTCAGGCGTCATAGATTG

>bc25mer_184954-20
AGAATGTGGAGGACATGTGG

>bc25mer_188166-20
TCGGACTGTGAGACATTTCC

>bc25mer_192833-20
ATCAAATGCCTCGGTCGTTG

>bc25mer_193757-20
GCAATGCTATGCTGGGACAA

>bc25mer_199694-20
ATGGAGGCCGTTCTCTGACA

>bc25mer_200273-20
GCAGATGGGTGCATTCATTC

>bc25mer_208214-20
TATGCACTCTGCGGATGACC

>bc25mer_218789-20
GACCCGGGAGTTATGATACG

>bc25mer_235702-20
CTCGGGTAGGCTATCTCCAA

>bc25mer_239358-20
CTGCAGGTGCTCTAGATGGA

>bc25mer_5878-20
GTCCAGGCTCGTCTTCTCGA

>bc25mer_9018-20
CTTACCTGAGCGCAGTTCGA

>bc25mer_9904-20
GTTGTTCGATCCCTCCACCA

>bc25mer_10475-20
CGGCAGAGGATAATCCTAGC

>bc25mer_11594-20
GGGAGTCCGGATGTTAGTGC

>bc25mer_12189-20
CGCATCTATGCTCCGCTTAC

>bc25mer_14363-20
AACGCGCTACCCAATTCTAG

>bc25mer_16178-20
TACTGGACGATTCCCGACTG

>bc25mer_17067-20
AGCGTCGTACCCAGTTAAGC

>bc25mer_18548-20
GAGATTCACGCCCTCATGAG

>bc25mer_20080-20
CCGCTAGCGTTACCTCTACC

>bc25mer_20601-20
CCACACCGTATGCATTTCTG

>bc25mer_26252-20
CCTTAGCGCACAAAGAGACG

>bc25mer_28353-20
GTCCTGATGGCTTTCTCACG

>bc25mer_30927-20
GGTCGTCGGTAGATCATTGC

FIG. 3-1

>bc25mer_32240-20
AGAGCCGGTATGATCCATCG

>bc25mer_32875-20
CCGGACGAAGGTTGATATCC

>bc25mer_39270-20
ACCCGGTACCTGTTATCACG

>bc25mer_40398-20
ATCCCGTCTCGTTCTATGGG

>bc25mer_41429-20
CAGGTGAGGGTCTCTCCTTC

>bc25mer_42283-20
CCTTATGGAAACGTGATGCC

>bc25mer_43506-20
GGCTGGGCATACATAGACCT

>bc25mer_45853-20
AGTGGCGGCTATTACCGGAT

>bc25mer_47706-20
TGATAACGTCCGCTCGTTGA

>bc25mer_54028-20
TACATCGTTGAGGCCCGTTT

>bc25mer_54772-20
CGTTGTGGCATCAGCTAGAG

>bc25mer_55112-20
ACTGGCCCACACACTTACCT

>bc25mer_57695-20
ATCGGGTCACGGAATATGAC

>bc25mer_58256-20
CGTAGGTTCACAGATTTGCG

>bc25mer_59008-20
TAGGGCACGAGAAGGTATCC

>bc25mer_61977-20
ACTAGCTTGTATCGCCGGAT

>bc25mer_63426-20
CGGCGGTTTGATATTCGAAG

>bc25mer_63840-20
GCCGAGTGTTTATGAGCAAG

>bc25mer_65782-20
TCTCCGAAGATCCGATATCG

>bc25mer_66426-20
GATGCAGGGCCTAATTAACG

>bc25mer_68963-20
TGACGCGGCTAAATACTGAC

>bc25mer_70828-20
GCAATGCACGCTCTCCTAGT

>bc25mer_71011-20
TGGCGTCAGATGGATTAGGA

>bc25mer_71626-20
GTTGGCGGGAGACTAAGAAG

>bc25mer_72701-20
TCGCGTCTCGTCCTTCTACC

>bc25mer_79345-20
CGGTCACGTCGGAAATAACC

>bc25mer_79641-20
ATTAACGGGCCAGGTTACTG

>bc25mer_81698-20
ATTGTATGGAGGCGCCCTAT

>bc25mer_82313-20
GCCAGCTTTCCAAGATTCAG

>bc25mer_83004-20
ACAACCGCGTGTTACAAGGC

>bc25mer_83517-20
GCTAATGTTTCCGGAGAAGC

>bc25mer_83687-20
GAAGTATCCGGCATCACAGC

>bc25mer_85062-20
TCGTTACGGACTTTCACGAC

>bc25mer_87436-20
GCCATTCTGACATACCCAGA

>bc25mer_90109-20
GTATTTGAACCGGCCAGCTG

>bc25mer_95782-20
GGTTGTGGGCCATATCCAAT

>bc25mer_96522-20
GGCGTTCCATCGAAACTCTA

>bc25mer_96896-20
GGCCATCGACTTAGATTCCA

>bc25mer_97519-20
GGTGCTCCTGCCATTATAGG

>bc25mer_100107-20
TGTCGACGTGCGGAAAGTAG

>bc25mer_100109-20
GCCGCACGAGTATGCTACTG

>bc25mer_100502-20
TGCTGCCGCTTACTACTGCT

>bc25mer_102626-20
TAGCACTCTTCTGGCCATCG

>bc25mer_103544-20
CGGTAATCGGTTCACCAGTG

>bc25mer_108258-20
TTTAGTGCCTGGCTCCGTTG

>bc25mer_109838-20
GGAGTTACAGCCACTTTCGG

>bc25mer_114408-20
GTGTGGACGGAACCTGACAA

>bc25mer_115589-20
CACACGTGGAATTGTTCTGC

>bc25mer_116534-20
CTTGTACATGGAGGGCGACT

>bc25mer_116605-20
AATCCGGTGTACAGGTTCCC

>bc25mer_121894-20
AGGTGCGACGATATACACCA

>bc25mer_123748-20
ACTGCATATGACCGCTGCAA

>bc25mer_125347-20
CCTTCCATTAGCCGAGATCA

>bc25mer_125541-20
GACGCGAGGAGTGATCGACT

FIG. 3-2

>bc25mer_126017-20
CGACCCTCACACATTTGGGT

>bc25mer_126439-20
CGCTCCCGTATTCTTCTGTG

>bc25mer_128324-20
CAGGGATAGTCAGGCCAGTG

>bc25mer_130576-20
CGTAATGAGTGCTTCGCCAT

>bc25mer_132438-20
CTCGTCTCGTGGAACACATG

>bc25mer_133420-20
TATGCTGGGCAGTAATCAGG

>bc25mer_133575-20
GCATCTCGTAGCATCCTGCT

>bc25mer_134218-20
AGTATGGTACCGGGAACAGC

>bc25mer_136989-20
ACCATTCTCGCAACTCGCTA

>bc25mer_137798-20
CGTGGGCGAAGTACTTGGTC

>bc25mer_138338-20
CCGGACGTCTTCGATAATGC

>bc25mer_139329-20
TGATGCTCTTTGCAGTTCGG

>bc25mer_139807-20
ATGCCGGTCATAACAGTGTC

>bc25mer_140760-20
TGGTGGCTCGTTATCACAAG

>bc25mer_141708-20
TCTGAAGCGTGGCCATTACC

>bc25mer_141799-20
CATATGCCGGACATTCAGCT

>bc25mer_142929-20
ACTAATGGTCCTGCGGCATA

>bc25mer_146048-20
GGTGCCGTGTTGCATGTAAG

>bc25mer_149885-20
AAGCGGTCGTGGTTTATACC

>bc25mer_150249-20
GCCACCTTGTATGGTATCGA

>bc25mer_150996-20
TAATGCTTAGGCCCGTCGGT

>bc25mer_153952-20
GACGCGACGGATTATTTAGG

>bc25mer_156264-20
GGTCGCGCCCATATATAAGG

>bc25mer_156364-20
TGACTACGGTTGGGTGCATC

>bc25mer_157007-20
TCAGGCGCTCATTGTATGTC

>bc25mer_160860-20
GTAGATACTCCCGGCCCGAA

>bc25mer_161471-20
GCAGCACTTAGGGCAGCATC

>bc25mer_163904-20
CGCGGGAACCATATTAGGAA

>bc25mer_164195-20
GATCCATGGCCAGTTCGTAT

>bc25mer_164787-20
GTGCGCCAAAGGACTTAGTG

>bc25mer_166664-20
GATCCCGCTATTCACCGATT

>bc25mer_167568-20
GCATAGATGGTTCACCCGTA

>bc25mer_169085-20
GTCTATAAGCCGCGCTGCAA

>bc25mer_170764-20
GCGGAGTATGCCATCATGAG

>bc25mer_171455-20
TGCCGCGATCATCTACTATG

>bc25mer_174112-20
GACCGGGTATTCGACGTCAT

>bc25mer_175501-20
GCAGTGCGGGTAGATACGCT

>bc25mer_180358-20
AGGCGTGGGTAGCAACGTAT

>bc25mer_180797-20
GCTCATTGGACTTTCTCCCA

>bc25mer_181432-20
AAGGACCACGTATCTGCATG

>bc25mer_181702-20
ATCTGATAGTGCCGCGACGT

>bc25mer_182418-20
TGACCTGCACGGATAGTAGG

>bc25mer_182774-20
GTTATTCGCAGTCCTTGGGT

>bc25mer_183321-20
GTGCAGTGTGCTTAAATCCG

>bc25mer_188636-20
CGCGGGTGTTAAATAAGGAC

>bc25mer_189039-20
GCGAGGCGTGGTAATAGTCA

>bc25mer_190422-20
CAAGGCGCAAACATAGACAG

>bc25mer_191525-20
GGAACAAGGGCGTCTATGTC

>bc25mer_193359-20
GCAGTGCGGATAAGCTACAC

>bc25mer_193734-20
AGAGATGCGTGTAGGCGATT

>bc25mer_194512-20
CCCTTTCGAGCTAAGTTTGG

>bc25mer_196660-20
CCTTAAGCAACCCGTCGATG

>bc25mer_197001-20
TTGCTCGTGACTGAACAACG

>bc25mer_197794-20
GGTCGTGCGTATAAGCCTCA

FIG. 3-3

>bc25mer_198508-20
CGGATGGTCTTCGTTTAACC

>bc25mer_198891-20
GGCACGGTGGCTAGTAACGA

>bc25mer_200581-20
GCCGGCCCAACTGATAGTAG

>bc25mer_200721-20
TCGACTTCGGTCACCTTTCC

>bc25mer_206382-20
CGGGTCGATACTTTCCTCGT

>bc25mer_207645-20
GTGCCGTGTGTAATATCCGA

>bc25mer_210956-20
CGCGCCTGGAATAACTCCTA

>bc25mer_210958-20
GTCAGGGAACCGTTTCTTCA

>bc25mer_210961-20
CTTGTCATGTACCCGAATGG

>bc25mer_211638-20
CGACCTGGCTACGTAGAACC

>bc25mer_211770-20
CTAGCTAAAGGGCCGTGCGT

>bc25mer_213351-20
AAGTATTAGCGCGGCAACGT

>bc25mer_213506-20
GGTTGCGTGCCACTTAAAGC

>bc25mer_216161-20
GGAGGTTCGGTTGTACTGCA

>bc25mer_216486-20
GTTGTCGTCCTCCATCGGTT

>bc25mer_218887-20
GGGACGGAACTACACATGTG

>bc25mer_221408-20
GATCCATCCTGATTGAGGGT

>bc25mer_222754-20
CTTCATGGTACCGGTTGAGA

>bc25mer_223329-20
TCGGTCGGCTGTAAGGATAC

>bc25mer_224422-20
ATTGCACGAGGTCAGAGTCG

>bc25mer_225143-20
GCTTAGATCCGCTCGCTACG

>bc25mer_227128-20
CGGCTGGCCTACTGTAGAGA

>bc25mer_227360-20
GCAACATGACCTGTCATCGC

>bc25mer_228435-20
GGCCGCACGATATATTTGAC

>bc25mer_229488-20
CAGCCGGGCCATAATAGTTG

>bc25mer_229709-20
TGCGGTGGACCTATTATCCT

>bc25mer_229929-20
CCTGCGTTAGGCAATCCATC

>bc25mer_230898-20
GTGGGCTCTTCGAAGTAACC

>bc25mer_233755-20
GCCGTGGACCACTAAAGTTC

>bc25mer_234401-20
GCTTAAGTCATGGGCGCATC

>bc25mer_236515-20
ACAGGTTAGTTCCGCGCACT

>bc25mer_237582-20
GGCGTGGCATTTAGACTACC

>bc25mer_238816-20
ACTACGGCCAACAACCAACA

>bc25mer_238839-20
GTTGACAAGGCTCTGTACGG

>bc25mer_239304-20
TTGTTCTCTTGCCGGTCGAT

>bc25mer_239849-20
GTTGATCGTAGCCAATTCGG

FIG. 3-4

Exp-1_acnB-73_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGCAGCTCTACAAGTGCGGC CATTTGGTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-191_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTTCGCGATAGCAGCCAGGA AGCCTGCTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-671_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGCGGGAAACCTTTCTGTT GCAGAGCTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2404_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCCAGTTCCGCAGAAGCCAG GAAGACATT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1307_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGTGTGGCAGAAAGACTGCA TCACCAGGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-779_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCCGCGTTTGTTCGGCACAT GTGGAATAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1343_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTGCGTGTTCACGTCAACTG GCTTCGGAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1961_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATCTGCATCGGCTTCCAGCA GCTCAGGAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2131_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCAGCACGGAAGTGACCGAT GTTGGTCAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2039_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATCCGGGTCGTTCGGAGCAC ACAGGATTG
CATTGC GCTTCTCCGGAAACATTAGC

FIG. 4-1

Exp-1_acnB-2262_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATACGCGCACCACTCTTACCG AAGACGCTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2542_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCGGCTTTCTCGGTGTACTG AGAAAGCTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2369_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGTACCCAGACGGTTCGGGA AGTTACGGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1811_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTTGTTCAGCTTGATGGTAC AACCAGCGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-428_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATCCGCCCAGGACTGCATAA CCTGCTTCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-744_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCCATAAACCACAGAACGGAG TTAGTGGCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-393_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCGTTGCCTGCTTTCGCTTTC TCTTCTACG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-530_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGGTGCCGGAGAAAGGTCAT CGGTGTTAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1378_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCGCCACGGTTCATAATGAA GTCCGGCAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2170_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTCGGCAACTGACCTTTATG CGCATCCAG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-2

Exp-1_acnB-943_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGTTCGCCGGTTTCGTGGTT ACGCACTTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1881_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCGCCGTAACCTTCCGCGATC ATCCACTTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-634_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTGATCGGACCAACAACACC AGGCTGGTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1224_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGGTGGTGTCCTGGGAACCT ACAGAAGTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1768_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGACGCATCGGTTAGCTCAAA GGCCTGCTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2473_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCTACCTGCGCCACGTAGGT CTGGTACTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1005_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATACGGCCACCAGCACGCACT TCATCAATC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-352_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCGAACATCAGCAGCGTGTG AGAAAGTGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2002_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AATATCCGCCAGATCGATGTC GATCACTGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2088_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AACACTTCGTCGATCTTCTCA CCCTGTACC CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-3

Exp-1_acnB-1681_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTGCCTTTCTTCTCAACGGT CAGCAGACC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1916_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCATGCCCTGAATACGACGTT CCAGGGTAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1478_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGAAACGGGTATGGGAGTCAC CACCGGTAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-706_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGAAGAACCCGTACCCACAAC GTCACCGAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-897_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGTCAATCACGTCGCCCATG TTCAGGTTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1515_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCAGACCAGAACCCGCCGGG AAAGAGATA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1581_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTGAAGCGCACCAGAACGGAT TCCGGCATA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2226_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTTCTTCGGTCAACTGTGCG GCGTCCATA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2297_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTTACCCATACACAGGGAAC AGCCAGGGA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1733_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCAGATCCGGCAGACCTTCAA TTTCCAGGA CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-4

Exp-1_acnB-305_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGTTTGGCATCATCCAGCG CGTCGATCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1266_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCCAGGCACGCCAGGTCTTTC AGTTCATCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1040_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTTGGTGGTCAGGCCACGCC CGATAATCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1646_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGATCGCATACAGCGGAATAG CGTGTACCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-848_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGTGCACCCGCGTCTTCCA TCGTGTTAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1436_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGCAGCATACGGTTCAGCC AGGAGTGAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1105_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTCTCAGCGACATCTTTCGC CTGACGGAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-646_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTTTCAAGGTCGCCCGCCAG TTTCTGTTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1951_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACGGTGCTGGAACCACCGGC ATAGAAGTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1354_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTCCCGTTGATACCAACAGC ATAACCTGT CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-5

Exp-1_bamA-686_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGAAACGGGCATAACCGCGAT CCAGATAGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1189_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGAACACGTTGGGTATCGGT ATCGACAGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1410_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGGTGAAGTACGGGTTGGTT ACCGACAGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-857_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACAGCTCACCCGGCTCGATCT TAGTCAGCT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1490_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGTTGGTATAGTCGGACAGGT CGGCGTCAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-85_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGCGACACGCTGAAGGCCTTC GAAATGAAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-564_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGGAAATGAGAGATCAGTTCG TCGGTGGTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-508_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTGCTGGATTTCAGCTGACAC ACCTTCCTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-967_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCGGCATCGTTAATTTCGGG CATCGACTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1989_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACTGCTTTCGGACCAATGGTA TTGGACTGG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-6

Exp-1_bamA-1224_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCTTGTAGACGACATCAACC TGGTCCGGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1787_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TATCCGATCCAGGAATGGTCA CTTTACCGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-50_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCACTACGAACCCTTCAGCAC CGTATACGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1101_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGCCATGCACCTTCCATCTGA CGCATTTCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1560_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCCAGACCTGCACGCAGCGAG TTATATTCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-2382_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AACTGGAACTGTTCTGCCTTG TCTCCATCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-145_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCTTCATCATTAACCGTGTC GCCTGTGCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1700_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCCAACCATAGTTGAACGTGA AGTCGTCCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-2299_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCATTGTAATGCGATACCCGC AGACATACG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-611_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATTTACGATCGCCTACCACGT TCCACCACG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-7

Exp-1_bamA-2229_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGGCTGGAATCCCAGTTTGTA TCCCAAACG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-423_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGCTGGCGCTATATTTACCG ACGCTGTAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-2095_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCGCCTACAGCATCATCCGA TTTACACAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-721_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCTGGCGTCAGACTGACCTG GGTAGAGTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-2059_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCGCCGTCCTGAGTCGCACA TTCGTAATC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1857_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACAACCCATTTGTGATCGTCA TCGATCGGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1630_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCCATAGAGTACAGATAACG CCACATCGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1260_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATACCAAAGTTGAAGCTACCG GTGTTGCGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1445_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TATAGAAGAGACGACCACCGA GGCTTACGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-892_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCATCTTCCATCTTGGTCAC TTTGGTGCC CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-8

Exp-1_bamA-1029_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGATCTTACGCACGTAGAAA CGGTTACCC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1066_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGGACGGCATCTTTCGAGGT ATCGTTACC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1154_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGAAGCCCAGACGATTCAGAC GCTCCTTAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1751_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGACACGTGAACCATCTGTCG GGAAGTAAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-15_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGCTAAACAGCAGCGACGCT ATGAGCAAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1595_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GAGGCTGCATGTTGGACAGGG AGTTATGTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1525_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGGAAGCCCAACGTCACGTC TGTACCATA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-2024_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGGATCATAATTACTGGCCT GATGCGGGA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1665_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AAGCTGTTATCCTGATCAGAG GTGCTCGGA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-234_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCTGAACCAGAAGGGTATCA CCATCACGA CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-9

Exp-1_bamA-1911_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGCATCTCTTTGCCGCCTAAA CCATCACCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1318_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGTTATCCTGCTGCACACC AGCCTGGAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-932_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGGATAGGCATAACCATAGC GACCGAGAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-803_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAAGGTTGCCGCTCACTTCAA CGCCAGAAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_bamA-199_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGAACATCCTCAAAGTTGCC GGTAGCAAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_btuF-721_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGTGGGCTTGCACGTTCAAA CCAGTCACT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_btuF-764_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTGTGAAAGCGCATTACAGA GCTGTTGTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_btuF-418_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTGCGCTTTCAATTGCGCGTA CTGATCCAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_btuF-143_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGCTTGTGGAGGATAGTCGG AATAGCTGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_btuF-637_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGGAATTTGGTCCGGTCCGCC TGTAATGAC CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-10

Exp-1_btuF-595_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGAGCGTGCTAACACCTGTTC GCGGCTAAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_btuF-240_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCATTACCTCCACGCCAGGCA ATCACCAGA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_btuF-87_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAAAGGCAAGTTCAGTGTTG GCGGGAGAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_can-349_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GATATCGCGGATATGCAGCAG CCAGTTGTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_can-202_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGTTCAGGTCAGTGTGAAT GACCAGGTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_can-167_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATTACGGTGAACAAAGAGTT CGCCCGGCT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_can-630_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTGTGGTTGGCGTGTTTCAGC TTGAGGTTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_can-132_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGGTTAAACGTTCTGCAGGA ACGCGACTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_can-434_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTGTTCCATGACGTTCAGTT CACACAAGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_can-237_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCGAGTACATCCACTGCATAC TGAACCACG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-11

Exp-1_can-480_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTTTCCACGCTGATTGCATA ATGGTGGAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_can-595_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCCGTGACGGTAACGTTGCTC AAGGGTTTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_can-279_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGCCGCAACCGTAGTGGCCA CAGATAATA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_can-559_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTGGCGGTAACATCCAGATC ACGCAGCAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-414_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAATGATTCTCGTCATAGTGC CAGGCCCGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-37_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGAAACAACGCCGCGATGAC GACGGGTAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-87_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCAGCATGCAGACCACCAGC GTTACAATG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-319_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGAGTTACGCCAGATTGCTGC GGAACCTGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-247_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGAAGCCCAAAGTGAGATTTC AACCAGCGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-809_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTAACAACAAGCAAGCAAAGA CCGGTACCG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-12

Exp-1_cdsA-449_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGATCATGACATAGAGCAGCC ATATTGCGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-749_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TACCACCGTGTCCTGGAATTA AATGACCGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-490_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTTGCCAAACATATATGCGCC GGAGTCAGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-575_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGCTGCAGTAGCGAGTCCAC CGATAAAGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-700_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGCTTAAACATACTCTCGGT CAGATCGCC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-179_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCATCAGCGCCAACAATAACC CGCATAACA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-618_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGACGTCGAGATTCGCCCAC ATGCCATAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-282_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGCACCAGCAATAGCGCGACA ATCCACCAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-665_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGAGGCTAACGCTGCGACAA TAGAGCAAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_dapD-653_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GAACGCGACCGTAGTGGATTT CGCCGGTTT CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-13

Exp-1_dapD-93_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGGAATCCAGCAGGGCGATC ACCTGATTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_dapD-384_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTGCCTTCATCAACATATGCG CCGATGTTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_dapD-148_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAACCACTGATGCGTCACCCA CTGACCGTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_dapD-51_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTAACGGTGTCTGCATTGGCT GGCGTGATC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_dapD-724_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AACCGCACAGTAGAGGCTGTA TTTGCCATC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_dapD-794_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGTCGATGGTACGCAGCAGTT CGTTAATGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_dapD-613_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AATACGGGTGCTCTGACCAAT GTATACGCC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_dapD-540_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACTTCAGAGCGCGCGCCGATG AAGCAATTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_dapD-348_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACGGCATCAGCACGGTGTTA CGGGCAATA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_dapD-257_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTCGTCGTAGTCGGCGAATT TCATCGGCA CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-14

Exp-1_dapD-292_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGGCACAACGCGGAAGCCTTC TTTCTGGAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_dapD-419_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACAAGAACCGACGGTCGCCC AGGTATCAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-302_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGGTAAACCGGCACCTTGCC ATAGAAGTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-95_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTCGACAGGATACGCGTCAA TACCGCCTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-814_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTGTGGCTGGTCATACCCGG TGAGGTATT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-766_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGTCCATCGCTGTCCAGCAT AACGTCAAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-614_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCGCCTCATAATCATAGAAGG TTCCGGACG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-726_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTTTGCAACCTAACGTCGTC CATGCTTTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-168_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGCGACCGTGTAGCGCGATA AACACTTTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-435_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACATTCCCACACTGGAACCT TCGCGGCTC CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-15

Exp-1_ddlB-559_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCTTCACCGAGTATCGCAAC CGTGAACTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-394_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AACCGGCAAACCCAGAGCAGA AATTTCTGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-691_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAATGCCTGCAAATTGGCCTC TTGTGACGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-259_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCCATTGAAAGCGCAGATGC CATCACTCC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-891_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAGTCCGCCAGTTCCAGAATT CGTACTACC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-133_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCCATCGACTTCAGTTGCGT CACGTCGAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-55_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGCTAACACCGCTGCGCCAGA ATTCAGAGA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-224_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGTATAAGGCAAGCCCATCA GCTCGAGCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-1219_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACGTTGTTCACTACCACGCC CTGATCTTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-703_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTTCAGGTTAACCAGCGCACC ACCGGAGTT CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-16

Exp-1_degP-1114_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGCAGTTCCAGGTTCACGTT AACCTGCTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-948_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGCGGAGGAATTAGGCAGA ACCTGGCTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-13_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACTCAGAGCCAGTGCACTCAG TGCTAATGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-1184_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGTTGCTCATCTCAGCGCCTT CAATGCCGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-1359_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGAATGTTGAGTGCCAGCACA GACGGTTTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-319_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCGGAACCCAGCGCCATGAA TTTCTGTTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-1149_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATGGAGCTGGAATCAACCTGA TTCTGGCTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-852_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATAATACCCAGCTCACCGCGT TTCACCTGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-1015_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGCAAAGCTGCTGATCGGCTT ACCGTTCAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-904_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGGGCGTCAACTTTCATCGC TTTCGCCAG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-17

Exp-1_degP-86_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCATCTGCTGGGCTGTCGTTG CTGAAGAAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-603_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGCGCAGAGACAATCCCGGAA GTTACCGTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-481_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGGTTCTGGATTTGGATCAG CGCGATATC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-741_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTGCGAGGATCGCGGTGTTG ATACCGATC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-1394_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACTGCATTAACAGGTAGATGG TGCTGTCGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-371_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAACAACGTGGTTGTTGGTGA CGACATAGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-411_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGCCATCGCTCAGTTGAACT TTAATGACC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-141_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGTTAATGCTGACCACTGAA GGCATCACC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-1294_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACTGCCTGCTGGTTCGCGCC AATAATCAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-528_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGCGCAGTGCATCAGAATCC GCCATCTTA CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-18

Exp-1_degP-565_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCAAACGGGTTACCAATCGC TACGGTGTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-177_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGCATACGCGGCGTATTAACG GTTGTGCTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-1076_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGTAAGCCCAGGGTCAGTT TGCTGCCTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-446_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGGATCTTTGCCAACCATCT TCGCGTCGA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-258_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGGCAGAACGGAGAGCTCTGG AACGGAGAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-223_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGGCAGAACGGAGAATCATC ACCGAAGAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_degP-48_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCGTTGCAGAGAGCGGAGAT AACGCCAAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_erpA-158_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGATGGTCATATCGCCTTCGT TCACCTGAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_erpA-63_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGATTCGGGTTATCTTCGTCA GCGATCAGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_erpA-314_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGATACTAAAGGAAGAACCGC AACCGCAGG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-19

Exp-1_erpA-28_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AACTTTGTTGGCTGCTGCGTC GGTAAACTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_erpA-238_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCTTCGGTATAATCAACGGA ACCGCCGAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_erpA-277_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCGTTCGGGTTGGTCACGAT GAAACGAGA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_erpA-123_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AAGGTGAAACCATACTGGAAG CCGCTGCAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-409_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCCTGCAACTTCAGGCCATT CAGATAGTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-599_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGTCTGGAACAGGCTGTCAG TACCGGCTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-529_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTGACCATATTCAACAGGCC GCCAGGACT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1514_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACCACCACGCCAGGTAAACT GTTTGTCAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1253_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTCGTAACCAAACCAGGCGT TGATGTCAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-490_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTAAAGCACGGAAACCGGGCC ACGCATAAT CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-20

Exp-1_fhuA-176_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGTAGCAGACTGTCGCGCCG CAATAGTTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1471_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTCCCGGCAACGCGGTTAAG AGATTCTTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1799_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGCTTTCGCTTCGATTTCTA CGCCACGTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1183_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCGATATCGCCAGTGGCAAA CTTGCTCTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1100_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TACGTGCCAGATAATGGCCTT TATCCGCTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1029_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCGGAGCAGACGCCGTAACCA TAAACGCTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1871_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGGTAGTATCGGTGGTGTATT CCGCATCGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1218_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATACGCATAAAGTCGACACCG GTCAGCAGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-264_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACTTCGGCTGATGCAGCGCC ATCTCTTCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1323_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCGGATCTTTGGCATTGAAG TCGAAATCG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-21

Exp-1_fhuA-2129_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACGATCGAACAGGTTGTTAA CATGCAGCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-873_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCATTAAAGTCTGTCGGCAGA CGCTTACCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-455_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCGTTCCAGCATATACGGGT CAATGACCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1693_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATACACGGCACCAGTAACTAC AATCGGACG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1435_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGTCATAACGACCGCCTAG GGTGACCAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1600_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCATCCTTCCCAACTTGCGA AGAAGGTTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-118_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCAGGTGCAGCGGTAACGGT GATAGTGTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-939_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AATTCGTGATCGAAGCTGTAG CCGACCATC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-2167_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GAAGCAGCCATAAGTGTTAAA GCAGCTGGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-2029_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCCCACTTTAAAGGAGTTAGC CGGATCACC CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-22

Exp-1_fhuA-2215_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTAGAAACGGAAGGTTGCGGT TGCAACGAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1288_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACCGGATTGTACAGATTGAG CAGTGGCAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-734_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCAGGTGAACGCCGGTGCAA TAGCATAAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-664_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCAAGACCGGTCAGGCGATA AGAGTAAAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1399_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCCCACTGCGCCTGATCCTG AACATAAAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-2084_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCATGCCGACTCGCGCCAGAT CATAACGTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-309_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTACCAACAGAGACACCCGGC GTGTAGCTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-358_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCAAAGCCGCGAATGATCAG GTGGTCATA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-564_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGAACTTCTTTCAGCGGTTCG GTGGTCGGA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-47_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCTAACCGCTGTGGCTACTA CAACTGCGA CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-23

Exp-1_fhuA-82_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGTTCAACCGCTGCCTGTGC ATAAACAGA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-796_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCGGTTTCCGGCTCGTTCTG GAAGTAAGA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-699_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCTTCTGACCCTTTCTGCTGG GCATTGGCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-980_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGCAAAGCGCAGGTTCTGAC GCACAGTAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1764_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCTCGCCACCTTCAACCGAG AAGAAGGAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1966_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGCGTCAGACCTGAAAGCGG ACCGTCAAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-257_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GAGGAAGCTGCTGCGGCAAAT AAGCCACTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-536_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGACAGCACATCAACCTGGT GGGCGATAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-133_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GAGCAGAGTGGATTTACCAGA ACCGTTGTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-720_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCCATCGGGATGCCATAAATC ATTTCGAGG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-24

Exp-1_fhuC-221_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAAACGCTTTGCTGCTCCAGC TTTCCAGCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-639_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGCGCAGGGCGACCAGATAA TCACAGTAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-310_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCACGGGTAACGACCAATCGC CACCAGTTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-500_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGAGGTCGGTTCGTCGAGCA ACAGACAAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-95_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACCGGTCACTTTCCCGGCAG GAAAGGTTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-578_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAATGACCGTCAGGCCACGCT CCTGACTTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-45_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTGCGCCCGGGCACACGAAAG GAGATATTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-431_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACGTTCGCCGCCAGAGAGAC TATCGACCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-680_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCATAATTTCCGCAGGCGTTC CCTGAGCAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_folK-321_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAGTGCGGAACGGTCAGGCGT TCAGTATTT CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-25

Exp-1_folK-16_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGAGAGGCCAGATTGCTGCC TATGGCAAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_folK-281_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TACCAAACAGCATGATGTCGA GATCCAGCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_folK-391_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCAGGAAACACCAACTCCGG CGCGATTTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_folK-148_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACGGCTGCGTTTAAGTAATC GGGTTGATC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_folK-80_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGTAAGAATGTGGCTTTCAG GGATATCGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_folK-183_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGTAGCTCTTCAGGTGCAAGA GAGGTTTCC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_folK-356_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGGCCACAGCATAAATCCAC GATTCTTCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_folK-230_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTTGCGGACGCGACCTTGCT GCAATTCAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-866_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTCGATCACCTCTGCCAGTG TCTGACGTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-608_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACCGATATCGACGACGCAGA CACCCAGTT CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-26

Exp-1_ftsA-721_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGCGTGCCAAAGGCGTAAGC GATATCACT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-43_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TACTAAAGCGGCAACCTTCGC GGTACCAAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-429_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCACGCCCGAAAGTCCTACC GGATTCTTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-475_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCCATATCGTTGTGACATGT GATCAGGTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-264_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGCACCATACCAATTTCATTC TGGCAGCTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-117_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTTTATCCATACCACGCGAC GGGCAGCTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-393_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCCTGATAGTCAATCGCATAC TCTTGCGGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-679_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCAGCATAAGGAATTACCTT AGTGTGGCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-1136_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACTCTTTCCCATAGTGAAGCA ATCCCACCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-1095_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TACGGCTCCTGAGCATAATCC GTTAAACCG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-27

Exp-1_ftsA-82_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCAATGATATTGACCATACC GTCGGGCAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-526_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGTTGGTCAACTTTCAGCCC ACAACGTTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-196_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGACAATCTGCCATCAATTC TGCCTGGTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-907_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TATCTCTTCGTTGACCAGGTT GAGCAGCTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-768_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCCAGCGCACAACCGTGGCGA ACTTTAATC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-562_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAATACCGAATAACTTGATGC CAGTCCGGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-1040_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCACTTGCGTATGAAACACGC GCTGAGCAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-1221_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCAGCCAACTATTGAGTCGC TTGATCCAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-998_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTCGATCTGCGCTGCGCCAC CGGTTAATA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-161_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCGTTGTACGCACTTGACCA CGGATTCGA CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-28

Exp-1_ftsQ-423_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTTCTGGCGGCACGCTGAAG GTATTTCCT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-464_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTCCGGGCCATACAGCATTG GAAGCACCT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-500_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTTCGCGATAGCCCTGCAACA CTTCATTGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-773_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTTGCTGAGTAGATTCCTCTG GCGGCAAGG CATTGC
GCTTCTCCGGAAACATTAGC
Exp-1_ftsQ-216_I1-bc25mer_83004-20_I2-bc25mer_83517-20

ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCCAATGCCAGGATCGACTGC CGGATATCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-290_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAATCCACGGCAGGCGTTGTT CTATTTGCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-251_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGTTGACATCCTGGGTCATAA AGGTACCCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-589_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCGTTATTCAGCGTCAACTG CCAGGAACG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-47_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGACGCGTTCCATTATTGC GGCGAGAAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-554_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGTCATCGCCGCTTCCTTCA GAGTAAATC CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-29

Exp-1_ftsQ-702_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGTAGCTAATCCGTTTGCCA TCGGTTTGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-388_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCTACCATATGTTGATCATT CCACCGCGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-329_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCAATTCATCAGGCCACTGCT TTCTGACGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-738_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGCCTACTGCCGCTCCAGAG TCATAACGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-181_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGTGTGTAATGGCGTTCACC GGTCAACAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-12_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCTTCTTCGCTGTTTCGCGTG TTCAGAGCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-829_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGCAAATGCACGGATGGTGTT ACCTACCGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-552_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGCACAGCGCCTTTCAGTACA TCGTTCGCT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-864_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGAGAAGTACCGATAACCACA GTCGCGTTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-969_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGCTGAACCTGCTTATTGGTC ACCAGAGTG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-30

Exp-1_ftsZ-899_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGTTACGCGCAGCTCGTCAT TCATATCCG CATTGC GCT-
TCTCCGGAAACATTAGC

Exp-1_ftsZ-505_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGGGAGATACCGCGGCCCAG AACTTTCAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-742_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCGCCAGACAGGTCGATATC TTCCAGCAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-781_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGGTCGAAGCCCGCCGTGAT GTTAACCAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-934_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGGACGTTTGTCCATGCCGAT ACCTGTCGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-706_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGAGATAGCCATTTCAGCAGC TTCTTCCGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-291_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCACCACCCATACCCGCAGCA ATAAAGACC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-326_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTCAGCGACGACTGGTGCTG CACCTGTAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-182_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGCGCCCAGTCCTTTGGTGA TACCGCTAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-1069_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCAGTTTGCGGCGCATTGTC ATTCACGAC CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-31

Exp-1_ftsZ-622_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATTACGGTGCGTACGTCTGC AAAGTCCAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-587_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCAAACCCGGACGAGTAATCA GTTCAGCGA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-377_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGAAAGGCTTAGTGACGACAG CAACGGTCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-27_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGACGCCGATGACTTTAATC ACCGCGTCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-1004_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCATCCCATGCTGCTGGTAGC GATCCATCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-470_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGTCGTTCGGGATAGTGATCA GAGAGTCCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-118_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGCGCTTGTGCATCGGTATT TACCGCGAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-77_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTCAATGCGCTCGCGCACCA TGTGTTCAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-965_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACCCGGTTTGGTGTCTGGCA TATTGCTTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-820_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATCACTTCCGGCGAAGCGGT TTCTGCTTT CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-32

Exp-1_gcd-1189_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGGTGCCCAGGAGTTTGGCGA GTTAAAGGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-785_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATGATAAGAGACACCACGGC AGGTTACGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1154_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGTGTTCGTCAGACGGGATTG CGTTCGGAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1502_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGGAACCACCAGTTCGCCAT TACGACGAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-368_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACCGCTAATCAGCAGTGCGA CCACCAGTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1447_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCCGGAGCGTAAATAACTGG CACTTTCTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-644_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCCCACTTTAATCGGCGTCA CTTCATTGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1975_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGCTGGCTGTTTACATGGCAG ACCAAATGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-877_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AATCGCAATCAGTCGACCATC ATTGACCGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-930_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTGAGCACGCCTTTATTGGCG AAGGTTTCG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-33

Exp-1_gcd-1616_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACATATCCGCACCGCTCAAAT CTTTCGTCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-542_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTTATCGGCGTTAATTTGTT TCAGCGGCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-23_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGGCTGTTAGCGTGACGAGTA ATCGTCGCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-2325_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAGTCGCCCATCTTCGTACCA AATGAACCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-126_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGAGCATCACAAGGCCAGCG ATAGGGTAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-403_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGCGGATCGTTAAATCCGGC CCAGGTCAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-679_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGGCGCTGGTGAGCGGTACA CAGGTAAAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1729_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTTACCCGGGAAGACCAGCGT ACCCTGTTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1000_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCGGTGATAATCGGTGGCGA AGTCGGTTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1387_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGTGCCGGAAGATCCATGTC CCACAGGTC CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-34

Exp-1_gcd-1240_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTCACGCCCATCGGCAGATA GACCAGATC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-728_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGGATCGTAATGCCATTTCT CTTTGCCGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-2362_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTACTTCACATCATCCGGCAG CGCATAAGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1352_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GAACGGTCTGGTAGCTCCACG CCAGTTTCC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-2290_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCTGCGGAGATCACCACATA CTGCTTACC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-2159_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGCCGTAGCGGCGATAAACA GCACGTTAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-2255_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTCATAGGTCATTGGCGTAG CCTGACCAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1789_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGCAATCGCCACTTCACGATT TGGATCAAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-247_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGAGTCAGCGCCCAGAAGTC GAAACCAAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1940_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGAACGGGTTGAGCGTGACAC CATACGGTA CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-35

Exp-1_gcd-2194_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCGTTGCTCATGTTGTAAGC GCGCAGGTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-603_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGATCGTTCGGCTGCTTCACA TCGCCAGTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-76_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACCAGCCAGCCTCCGCCAAT GAGTAGATA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1682_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGCCTTCATAGCGCATCTGGT GGAACATCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-507_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTTGACCTTCCTGATTACGA CCATAGGCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-2048_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTGCGGCGTACCAATACGTT TCTTCCACA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1839_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGGACCACGCGGGATCAGT TTCGAAACA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-2113_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCCGCCCAGCATCGGCATACC CATATTGAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-298_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACAAACGGCAGGATCAGCCA GATGCCGAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1066_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GATCACGCCAGACGTTTCGCG GGTTGAGAA CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-36

Exp-1_guaC-112_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACGCCGGACCAGCTCTGACC TGAATGTTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_guaC-407_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGCTTTCGCAACGAACTGCA CGAAGTGTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_guaC-148_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGTGCCTACGGTGTCCATATT TGCGGCGAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_guaC-458_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACCAGTCACTACGTTACCAG CACAAATGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_guaC-351_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAAACGAAGTTTAATGCCGGG TTCAGGTCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_guaC-760_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AAATTTCTCGCCGTTCTCCTC AACGATGCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_guaC-39_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGAGTGGAGCGTTTAGGGCGG ATGAGAACG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_guaC-641_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGGTGCAGCCACCATCGCTGA CGATCATTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_guaC-499_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AACGATATCGGCACCTGAGAG GATAAGCTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_guaC-989_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTGTTCCTGCACACGAATAA ACGTGGTGC CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-37

Exp-1_guaC-74_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATTGACGTTCCAGTTCAACAT CGGAACGGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_guaC-541_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACGCGAGTTGTACAAACAGA ACCTGGGCC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_guaC-811_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCAACGTGACGTTTCATCGC AGACTCGGA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_guaC-598_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCAGCATCGGCACATTCGAT TACCGCAGA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_guaC-302_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGAAATCCGCATCAGACGTAC CGGTAGAAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hemL-416_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGCAGTCAGCGTGACCATGGT AACACCCTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hemL-298_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACCATATCCATGGTCGGGAC CAGTTCGGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hemL-996_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTTCTTCTGCCGCTTCCAGC AGACCTTCT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hemL-250_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACTTCCATTTCGGTTGGTGC ACCAAAGCT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hemL-628_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GAACTCTGGCAGCGGCGGAAC ACAGTTCAT CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-38

Exp-1_hemL-961_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGTGTTGTCAGCTCATCCAG CGTTTCGTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hemL-1211_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TATTGATATCTTCCATGCTGT GCGCCACGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hemL-708_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTACGCGGAAACCGGTCATC ACTTCATCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hemL-120_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTGCCATCAACATCGTACAGA TAAGCGCCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hemL-762_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGAGGCAGGTTAAATCTGGC ACTACGCCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hemL-547_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTCAAATGCGGCGCGTACAGA AGCCAGATC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hemL-1249_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAACTTCGCAAACACCCGACG TGCAGCATC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hemL-1113_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCTTAAAGCGTTCCACGTCA CAGGCCATC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hemL-500_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGGTATATTTGGCGAAATCTG CCGGAACGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hemL-333_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTCATGGTCGCTTCAGTGCCG GAGTTCACC CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-39

Exp-1_hemL-830_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGCGCATCCATTACATCAC GACGACCAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hemL-1171_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCTTCAAACGCTGACGGTGC CAGGTAAAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hemL-923_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTGCGCGACTTCATTCAGAC AGGCGAAAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hemL-888_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCATCGCAATCGGGTTACCG GAAAGCGTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hemL-1040_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TACCGAACATGCCGCCAACGT GGTTAACGA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hemL-452_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTAACGTGAGTGCGCCAGAAC CGGCTTTCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hemL-593_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGCCACCGGCTCGACGATAA TACAGGCAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hpt-483_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGATATACGGCAGATGACGG TAACGCTGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hpt-80_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCATATCGCTGCCGCTGTCTT TGTAACGCT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hpt-330_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTTCGCGCAGGCTTAAGATC TCACGCACT CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-40

Exp-1_hpt-153_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCATGAGATACCTGAACTTCA CGGCACAGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hpt-365_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTTTATCCAGCAGCGTACAAA TCGCCAGCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hpt-400_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AAATTCTACCGGGACGTTCAC TTCACGACG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hpt-271_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCTTCAACAATCAGCACGTC CTTGCCACG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hpt-189_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGCTACCGTAGCTGGAGGCG GTCATAAAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hpt-447_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCAATGCCGTAACCCACCACA AACTCATCC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hpt-115_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AAACATAAATGAGCCACGCAG CAGACCCAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_hpt-41_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACGACCCAGTTCGGCGATAC GCGCTTTAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ispU-376_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTATTCCCGGCTGTTAGCGC TTCAGATTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ispU-339_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTTCTTGCAAACGCGAGTTA AAGCGACTG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-41

Exp-1_ispU-304_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCAATAATACGCAGACGCAC GTTATGTCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ispU-180_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AAGGCATACAGCGTTAACGCC TCAATACCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ispU-411_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCACCGTAGTTCGCCGCAATA TTCAGCGTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ispU-486_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCTGATCTGGTTGCAGGTTT CCTTGCTGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ispU-106_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGCTTTATGCCCAAAGGCACG AATCTTCCC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ispU-446_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGTTGCCTGACTCCCTGGA CTATATCCC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ispU-655_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCTTGTTCATCGAAATCGGG CCAGAGAAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ispU-257_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCTATCGAGCGCCCACACAA ACAGTTCCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_ispU-706_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCGCCGAAACGACGCTCTCG ATTAGCAAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_map-389_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCATGCGTAGCGCCAGGTACA GGCTTTCTT CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-42

Exp-1_map-151_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GAGGCAGGCAGAAACCGCGTG TTGTTCATT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_map-301_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGAGGTATCGCCGTGGAAACC ATCTTTGAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_map-766_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTATTCGTCGTGCGAGATTAT CGCCGGGAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_map-79_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACGCCCGGTTTAACATACGG TTCGATCAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_map-187_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGAGATGCAAACGGATTTCGG ATAGCCGTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_map-114_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAATCATTACAGATGCGATCC AGCTCGCCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_map-247_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCGCCATCTTTCAGCAGCTT AGCATCGTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_map-597_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGTTGACCATTGGCTCGATG GTGAACGTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_map-639_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCGTCCAGCCATCTTTCATG GTGCGGATC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_map-731_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATCCTTGCGTAGCGTCAGAA TTTCGCAGC CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-43

Exp-1_map-481_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCCGCAATATTCACGAACGAC GGAGAAGCC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_map-516_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGCGGTTCTTCATGGAAGCCG CGACCAATA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_map-562_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGGTTTCAGTACGACGTTGGT TTCACGGGA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_map-446_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTCGACAAATTTCTGAATCG CCGCACCGA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_map-338_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCCCATGATGGTCGGCTTAC CGACGATAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1945_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCCTGCGGGAAGCTCTGATA CAGCACTTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-146_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTTACGCCCTTTGCCTTTGC CCTTACCTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1588_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCAATTGGCGCATCCGCAAT CCACGTATT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1167_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGCGGACGGGCACTCAACATG TCATAGAGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1001_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGCTTAGCTCTTCTACCGGGC GACCAAAGT CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-44

Exp-1_mrcB-861_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATGATCAGCGCCATGTAAGCT TCGTTCGCT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1670_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCACCAGCATCACTCTGCCGC TTTCGCTAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-944_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGATTTCGTTGTCGCCGCTCT GACCGAGAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-2181_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGGCACCATACAGTTTGGTC GGCTGGTTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1036_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTTCACCATACCGACTAACAG CGCCTGCTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-446_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGGACGGCGAATCATCTCAA TGCTGTTGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1635_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGTCATCATTCTGCGGTGAC CAGACCTGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1710_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGATTTACCGTCGGCACGTTC ATCGAACGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-325_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTCATGTCTGGCTCAAGATT GACCATTCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1457_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAAACTGCGGCTCAGAACCTC CGACCATCG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-45

Exp-1_mrcB-896_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTCAAGAATACGGTCTTTGC TGTAACGCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-2493_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AACATATCCTTGATCCAACCG GCTACACCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-826_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGTAAGAACGCTCGCTGGA GAGGAACAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-582_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATGGTGATCAGACGCGGATCA AGACGGAAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-617_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAAACAGACGCTGCTCACCGT TTGGCGAAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-363_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCGCCTCCAGCAGCTTCACC ATCTCGTTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-19_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTCGGTTTCCCTTTGCGTCC AATTGGCTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-2425_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGCTGTTGCTGCGGCTGAGA AGACTGATC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-541_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGGTTGTTCTCCATATTGAC GATCGTCGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1543_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGCTGGCTTAAGGCCGTCAG ATAAGTCGC CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-46

Exp-1_mrcB-234_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCGAGATAAACGCCGTAAATG GCGATCAGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1128_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATAATCTGTTGCTGTTGCAGC AGACGCAGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-678_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGACGGTCTTCTGTCGCCAGC AAAGTATCC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1377_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCGCTCAACTTACGCTGTTTC TTCAGTGCC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-784_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGTTGCGTCAGCGTACTCGC ACCCTGTAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-2244_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACAAGATTCAGCGGCGTTGGC GTCTGGTTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-736_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTTTGCCAGCACCGCACGTCC GATTGAGTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-2117_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGCCGTCAATGCCCGCAAACC AGGTATCTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-2012_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCGTTGTACCACCTGCTGCA TGGTCCATA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-282_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGGCAGTTGCCAGACCTTG CCATCAATA CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-47

Exp-1_mrcB-2286_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGTCGTAGTCCACGCCCATA TCTGCAATA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1422_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACTTCACCACTAAAGCGGTCG ACGACCACA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1766_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTACGCCCAGTTTAATCCAGG TCTCCGTAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1241_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTGCAGCTCCTGACGCACCA GTTGCATAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1315_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCCTGGGCCACCGAGTCAAA GGTAGTGAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-481_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGCACCTGTCCTTCTTTACT GTCCGGGAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-2321_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAAGATACGCATGCCACCGC TGCAAACAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mtn-253_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCGTCCGAGACAACGATATC GCCCACTTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mtn-323_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGCCGGACAGCCTGGTAACT GACCGTATT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mtn-366_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGGCCTCAGCGGCAGCGATC AGTTTATCG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-48

Exp-1_mtn-67_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCGCAACCGCCGAGACTGAT AGTTTGACG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mtn-634_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATCAGGCTGGACTGTTTAGC GGCAACAGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mtn-183_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCACATCTGGCTTGCAGTGT TCCAACAGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mtn-448_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCCAGACCAACAGAACCGTT GATGAAAGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mtn-594_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCGAAGCTAAGATGAGACTGT TGATCGGCC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mtn-669_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAGCCATGTGCAAGTTTCTGC ACCAGTGAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mtn-559_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGAGATGGCGCGTACGACAAC AAACGGGAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mtn-102_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAACCTCGGTTCCATTCAGT TGGCCGGTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mtn-512_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATGGGCGATTGCCGTCGCTT CCATCTCTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mtn-137_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGCAGCGACTTTACCGATGC CCGATTTCA CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-49

Exp-1_mtn-413_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCTAACAATCAGGCCACGTA CAGCGTTAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_mtn-288_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAAATGCCGTGACATCCGCG TCGTGATAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_murC-1430_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTCCGGAGTTTGCGGCTTCA GTTTGATTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_murC-282_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGAATACGCGCTTCATGAGCG GCGACAATT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_murC-1064_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTTAATGGTGGCGTCCACTT CCGTCGGGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_murC-698_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTCGCGGTAACAATTCGCGGA TCACCGGAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_murC-741_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCGGCATCTTCGCTGAAGCCG TAAGTCGTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_murC-550_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTGACAATCGCCACCATCGG TTGCAGATG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_murC-597_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCTCAAAGTCGCCCTGGTAG GTATCCATG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_murC-402_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTTCTGCGTAGATGCTGGAA ACCATCGCG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-50

Exp-1_murC-911_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAGCCTCGTCGTCAATGCCCT CTTCCGTAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_murC-156_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCATTAACTGCTGCGTGACC GGATTTGGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_murC-106_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGATAACCTTCATTGGCCAG AACTTCGGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_murC-994_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCATTCACTGGCTCCAGCGG GAATTCACC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_murC-191_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGGGCGATGGTTGAAATAAA TCGTCGCAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_murC-776_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGGGCCAATCTGCTGATAAT CTTCTACAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_murC-1268_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACGTCCACGAATTGTGCGAC ACAGCGAAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_murC-247_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTTATCGGCAGAAATCGCGCT GGAAACAAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_murC-1337_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGTTAATACCGGTGCCAGCA TCTCGGCTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_murC-321_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCATTAACTCAGCCAGCATT TCGGCACGA CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-51

Exp-1_murC-1214_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTCGCCAGCCGGATACACTT CCAGCATCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_murC-515_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGATGCATCACTCTCATCTG CTTCGGCAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_murC-1179_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTATCAACCTGCGTCAGCACA TTGGCGAAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panB-616_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATCACGAGGATCTGCCCGTC AGTGACGTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panB-578_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAATGCCAATAACCGGGATCG CCAGTGCTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panB-304_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACCATGTTAGCACCGGCACG CATAACCGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panB-62_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGGCGAAGCTATAGTCATAAG CGGTGATGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panB-653_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAGGAATGTGACCGCCGGTAA TACCAAAGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panB-8_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGTACTTCTGCAGTAAGGAGA TGGTGGTCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panB-762_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGGAAACTGTGTTCTTCGCCC GGATAAACG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-52

Exp-1_panB-490_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGCAGCTTCTAAGGCTAATGC ATCGCTGAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panB-727_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTCCACTTCAGCCATATACTG CCGCACAGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panB-386_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AACCTAAGTGACCACATACAG GAACGGCAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panB-351_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCAGCATTTGTACGGTTTCT ACCAGCCAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panB-180_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGTGTGGTAGGCGATATCG GCAACGGTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panB-242_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATACGCCATAAACGGCAGGT CAGCCAGCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panB-113_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGCGAATCGCCCACCAGCA TGACGTTAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panC-775_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGGCGGCTACCAGAATTAC TGCCCGTTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panC-406_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GATGTCCGGCTGGACCAGGTT GAACAGCTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panC-37_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTGCCTTCCATACGCAGGCG GCGAATTTG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-53

Exp-1_panC-226_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGTTTGTTTAGCTTCTCGCA GTCCTCCTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panC-297_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGTAAGTGTGGGTTTCAGTA CCGTTCGGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panC-565_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGTTGTTCCGCCGTCAGATA ACCGTTACG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panC-76_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCGTGCAGGTTACCCATGGT AGGCACCAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panC-523_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGACCGTCTTTGGCGCGCAT AATTGGCAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panC-111_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGGCTTTGGCTTCGTCGACC AGCTTCATA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panC-2_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGCAGCGGCAGGGTTTCGA TAATTAACA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panC-332_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACCTTCCAGCATGGTCGAAA GGCCAGGAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panD-26_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGCATGAGTCACTTTCACGC GGTGGAGTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panD-274_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGGTGCGAGCTTCTTCATC TGGCATGGT CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-54

Exp-1_panD-156_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCGCGATGGCATAAGTGGAG AAACGCTTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panD-63_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCCTGGTCAATGGCGCAAGAA CCTTCATAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panD-352_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCAAGCAACCTGTACCGGAAT CGCTTTCGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panD-239_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGCTGGCGATGATGACAATAT CGCCGACAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_panD-204_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGTGGGCCGCCGCACCGTTA ACAGAAATA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-1007_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGAAAGCGTCGTGATAGGTCA GGCCGCTTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-652_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTTTCCGGGCTGATGCGCAT ACCCAATTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-1102_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGCAACTGCCAGATATCGCG GGTTAATGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-920_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGAACGCCGGGTTCACGCGCA TATCGTTAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-1314_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCGGTGACGGTTCTTCATCC AGCTCGTTG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-55

Exp-1_pcnB-960_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCGTCTCCAGCAGTGGGTAC CAGAACATG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-478_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATACAGGCTGTTGATAGTGAA ATCGCGGCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-728_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTTTAAGCGATTCTTCAAACA GGCGTGCCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-767_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACAGCTTATAGGTTTCGTAAC CGTAGCCCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-268_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGTTACGGAACAGTTTGCG CACCTGCTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-569_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTTCCGGGTTACCAATCAGAC GGATAACGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-1176_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCATAAGCCGCACGGAACTTA GGATGCTCC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-533_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCAGATCCTTCATGCCGCCAA CGTAATCAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-371_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTCGCTGACGTTACCTTCGT GGTGTCCAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-336_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCGCAACTTCGATAATCTCC GGGCCAAAC CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-56

Exp-1_pcnB-168_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAACCAGCCAGGCTTCGTAT CCCGCTTTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-604_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGCAGCATACGTACCGGATC TTCACGGTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-1137_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATGCGCGTTTACCCTGACGA CGGGACATA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-885_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCGTATCGGTATTCTTCAGC ACCTGTTCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-1211_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTTACGCTCAACTTCAGCTC GCAAGGCCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-824_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGTGAAGTAGCGGGTAATGG TCGGGAACA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pyrH-497_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGGATCAGCGGTAAACACGC CGTCAACTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pyrH-350_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAACAGGCTGATAGCTTCTG CCCAGCTGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pyrH-551_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGCACTTCGCTGTAAGTCA GTTGCTCGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pyrH-462_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTTTCAGCACCACATCGGCT TCAATTTCG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-57

Exp-1_pyrH-139_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCAATCACCACACCAACCTG AATACCCAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pyrH-91_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCCTGAGCCATACGATCCAG TATGCTTGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pyrH-618_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGAATCGGTAATTTATGGTCA CGAGCCAGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pyrH-315_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCGCACACGCCATTCAATGGA ATAGCGGAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pyrH-42_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTGCCCTGCAGAGCTTCGCCA CTCAACTTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pyrH-425_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGGCAAGCTGCTGAGTCGG TGGTAAAGA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_pyrH-390_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTACCTGTACCGGCGGAGAGG ATCACCACA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_rseP-723_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCATCGACTTTAACGATCCTG TCGCCTGCT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_rseP-959_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTCGACGATGGCGTTGAACG GCCCATACT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_rseP-1068_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTGGCGATAGAGATCGGCCCA CTGAGGTTG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-58

Exp-1_rseP-565_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGTAAATCGAGCTTTACATC CCGCCGTTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_rseP-64_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGGGCAACCCAGAAATGACC AAATTCATG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_rseP-758_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCATCACAAAGGTCACCCACT GCGTTAAGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_rseP-245_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TATTATTGAAGGCATGGTGGC GGAGTTCCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_rseP-866_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCGCTTTACCATTACCCGGTT TACTCTCCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_rseP-793_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAACGCTAAGGATTTACCCGG GTTATCCCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_rseP-650_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGTACAGGTTCAATTTGCG GCCCACGAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_rseP-1138_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTTCACGCTAATAAGCGCAAG AAACGGCAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_rseP-172_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCCAACGGGATCAGGGCGAT AACATATTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_rseP-913_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTCATCTGGCAAAGGAATGAC TTTCGGCTC CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-59

Exp-1_rseP-481_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACCAACTGCAAACGCACGGC ATCCCAATC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_rseP-530_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGCCAAATGGCGCTACTGTAA TGGTGGTGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_rseP-1286_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTAACAGCACCAGCAGAATCG AGCCAATGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_rseP-1008_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCCAGCATACTGACCGTCAGC TTCATCAGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_rseP-25_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AAGTACACCCAGTGCAACGAT GAACGAAGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_rseP-423_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGTTCCGTACCTGGTGCAATT TGTGCTTCC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_rseP-600_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCGGATCTTCTTTATCAGGC TCAAACGCC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_rseP-387_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCGAATTGGCTGCTATTTCG CCAACCACC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_rseP-134_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTTATCAGTTCGCCGCCAGA GCGCCTTAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_rseP-99_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTATTGAGAAACGCTCAACG CGAACACCA CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-60

Exp-1_rseP-299_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGAAGTTTGCAACCGGACCTG CGGCAATAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-907_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCCGTTACGTGGTGCATCAG CATGATGTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-1073_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTTCTGGATCTGCACACCTT CTTTCGCTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-77_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCATCTCCGGTTCCATGGCAT TGATGATGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-1859_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTTGGCAATCGCTTTAGTCA CCCACGGGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-1760_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TACGCATCAGCGCATCTTCCA TCGACAGGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-2308_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGGTGCTCTTTCCACAGGGA GTCAAGCGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-2184_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCTCACGCAGCGTCTCTTCA TGCAGTTCT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-1703_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACGACCAGAACGACCGCGCA ACTGGTTAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-1535_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGCCTGCCAGCTACCACCGA GCACAATAT CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-61

Exp-1_secA-1666_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGGGATTCGTGACGCTCGGT ACCGATGAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-2631_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACGGGCAAGGATCGTTACGT CCTACTTTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-2078_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TATCCCACATTTCTTCCAGCG ACTGTGGTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-610_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGAGTCCACTTCGTCCACCAG CGCATAGTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-247_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATACCGCCGAGTAACTGAAC GTCGAAGTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-1348_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GATGGAGATAGTACCCACCAG CACCGGCTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-2427_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCAACGACTCCAGCATCGCT GCAAACATG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-681_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTATACATTTCCGAGCTGTCT TCTGCCGGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-351_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGCACGCCTTTACCGGTTAGT GCGTTCAGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-2500_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGTTGTTCCAGCTCCTCAAC CTCTTCAGG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-62

Exp-1_secA-540_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTGTCGCGCAGGTAGTCAAAG CCGTATTCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-1121_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACAGACGGAAGTAGTTCTGGA AGGTGATCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-2148_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTATCCAGCCACTCGGCAATT GGCAAATCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-575_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TACGCTGTACACGTTCTTCAG GGCTGAACG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-1435_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCTTCGTTGGCGTGGAATTT GGCGTTCAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-454_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATGCCCGGCAGGTTGATACC GACAGTCAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-1267_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCTTCAGTCATGTAGACCAG GTCCGGCAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-853_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCATCCATGATGCCCTCTTT CACCAGCAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-955_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AACGATGTAGTCGACGTCACG GGTAAACAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-763_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGAGAAGTGGCCTTCGCCCTG GAAGGTTTC CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-63

Exp-1_secA-1500_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCACGACCCGCCATATTGGTC GCGATAGTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-2343_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGGTGGATACCCTGACGCAGA TAGTCCATC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-1907_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGTTGCTTACGAATGTCGA AGTTACGGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-2221_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTTACGCTGATACACTTCGAT GGACTGCGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-1970_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACAGTTCGTTACGCTGGGAGT AAATGGCGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-1821_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTTCGCCTGGCTTCATACCC AGTTTACGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-505_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTACCGTAAGTGATGTCAGC TGCGTAAGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-1626_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTGCTTCCAGTACCGCATCG TGACGTACC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-1392_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATACCGGCTTTGGTCAGTTCG TTTGACACC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-2005_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTGTTAATGGTTTCGCTCAC ATCGCTGAC CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-64

Exp-1_secA-1231_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGAATCATTGGACGGTTGGT CGGAACAAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-282_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTACGCATTTCGGCGATGCAG CGTTCGTTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-212_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TACCAAAGACGCGCTTACTTG CCTCACGTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-2535_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGCATTTGCGCTAAACGCTCG GCTTCCATA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-2273_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCATGACGCCTTTCTCGAAGT GACGCATCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-386_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACGTTGCGCCAGGTAGTCGT TGACGGTAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-2570_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGCTGCAGAGTCGTCATCCT GATGGCTAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secA-1007_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTGCATGGTACGACCGGTGT GTTCGTCAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secM-45_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCATCCCTAATAAGAGATGC GGCCAGAAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secM-346_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGCGCGCTGAGCGTATCCAG TAATGCAAG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-65

Exp-1_secM-174_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCGCTTCCAGCAAGGCCAAT TGACCAAAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secM-81_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGTTGCTGAGCGCAGGCAAA CCTAAACTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secM-209_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AATCAACGGAATAGTTCGAAT TCGGGCGGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_secM-311_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGCCTGAAGAGGCAAAGATT CTTCAGCAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speD-215_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGAATAGTGACGCTGGCAC CCTGTGGTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speD-407_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGGAGAAATCACGCCGCAGG TAGAGACTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speD-689_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCTGTCGGTTAAGTCTTCCG GTTTGGTGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speD-724_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCTTTCCACAGCGCAGCGGT AATTTCCTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speD-759_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTGGCATATTGCGCCCGTAA TAAATCTCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speD-593_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGACGTTCACATCCACCATGT CATACAGCG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-66

Exp-1_speD-139_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACAGGTTTCTGACAGGATTTC GGTCAGACG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speD-371_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGCGCGGAAGGTACATAAAC CGCCTTCAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speD-301_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCAAGATGGGCAACGACCGT TTCTGGCAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speD-250_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCGATGAGTTTCGGGTCAAC CGGTTCTTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speD-85_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATAAGCAATATAACCGTCGCG CTCTTCGGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speD-442_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTCAAGCTGGTGGATCAGGTA ATTCAGCGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speD-477_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGCGCACGCGATAATCAATG GTTACGATA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speE-622_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCAACGTCGCTGAAGTAATG GCTGAGTTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speE-367_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGATCGTCGTAGCTACCGGC GTTATGGTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speE-159_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAGATAAACTCGTCGCGCTCG GTGGTTTGT CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-67

Exp-1_speE-194_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGGCCAGTAGCGGAACATGGG TCATCATCT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speE-229_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCGCCGATAATCAGCACATG TTTCGCGTG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speE-462_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGATAGGATCGGTGCAGTCG GAGATAATG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speE-729_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCGCCTGAATAATTTCGGTT GAGAGATGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speE-783_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCGTATGGATTGCCGGATTG TAATAACGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speE-332_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GATACTGACGGCAGAACGATA CGACACCCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speE-122_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATCCAGCGCCATTACGCGAC CAAATGCAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speE-406_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AACGAAATTGACGCCATCGTC GATCACCAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speE-37_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCTACCGCAAAGTACTGCCC AAACTGGTC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speE-666_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCATGATACCGCCGTAATAG GTCGGGATC CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-68

Exp-1_speE-542_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTGCGACGAAGATACCGCCAG GATTCAGGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_speE-824_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AAGCCAGTGCGTCTTGCAGAT ACTGAGGTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacC-248_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAATGCCGCTGAGATCGCGGT AGCTGTCTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacC-80_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGCCGTTAAATCGGCCATAT CTTCGGCTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacC-21_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGACGGCCATCAGGCTGCCG AATAACACT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacC-315_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTGACGTAGGCAAGCAGGCTT AAGGAATCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacC-203_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGTCGAAGGTGTCGTAATTAC TGAGGTCCC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacC-149_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGTGCGCGACGAATTTGCC CGTTAGGTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacF-506_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGACTGGCGAATTAAATCCA GCACCATGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacF-117_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCACTGACATTACGGAAGAAA TGCAGCGCG CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-69

Exp-1_yacF-715_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTAGCAACAGGCCAGTTCGAA ATCCAGACG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacF-613_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTGCGGATAAAGCTGTGAATC GAGCGACAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacF-210_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGCCAATCCAGGTCTGGAGT TTACGTTGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacF-245_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTAATGCTTCAATACGGCTCT GGTCCACGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacF-385_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGTAGGTAAATCAAAGCTGCA ACAGCCGCC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacF-175_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCAAGTTCTTTCAACAGCTC AGTGCGGAC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacF-305_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GAAATTGCCCGATACGCGGCG CGGAAATTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacF-461_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGGGTTCAGGCTGGCAATCC AGGTTTCTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacF-349_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTCAGTCGCTGACGCACCAG AGCAATCAA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacG-10_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCCGCAGGTTGGGCAATTCAC CGTAATAGT CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-70

Exp-1_yacG-169_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCACTGCTTTGGTTCTTCGCT CCAGTCATC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacG-101_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTCAGCAGCCCATTCTCCGA GGTCGATCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacL-284_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTCAACGCCGCATAGCGACA GGCTTTCTT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacL-191_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCATCACCTCTTCACCGTCCA TCCACAGGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacL-110_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGCTGCTTGTTCCACTTCAT CAAGCAAGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacL-145_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGTTGCCAGGACCGTTCGCT ACCTTTCAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacL-226_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCATCGCCAGCGAATTCCAG TTGATTGGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacL-333_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TACTTCTGCTGCACGAAATTG CGGTAAGCC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacL-69_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCACCTCTTCATTAAACCAG TGCCCGACC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yacL-27_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATGGACATACGCACCTTTACC ACTCCGGTA CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-71

Exp-1_yadG-726_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGTCGATACTGATAGCCATCG AGCTTCGGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadG-761_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGCACTTCAACTTCCAGCG TCGCGGTAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadG-575_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCATTTCTGCTTCTTCCAGGT AGTGTGTGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadG-870_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCAGTGAAACAAACAGCTCT TCCAGACGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadG-281_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGCCTGATTCACCACAATTT GCTGCACGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadG-393_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCCCGCCAGATAACATACGC GCACGTTCG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadG-227_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTGCGGCACCAGTCCCAACT GACGTTTAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadG-428_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCATTAACGCACGGGCAATCA TTAAACGGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadG-497_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACATTGAGCGGCGAAGTTCAA TATCCACGC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadG-182_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCTCGAGATCGTAACCAAATA CGCTGACCC CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-72

Exp-1_yadG-316_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GATGTACGCTTCTTTGCGCTC CACGCCGTA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadG-685_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGTGCGAGATCGAGAATAAA GGTTTCCGA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadG-617_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCACCAGCTCACCGTGTTGAA TAATGCCGA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadG-66_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCACCCGCTTCGACCTGCAAA TCTATCCCA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadH-217_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGATGACGCAACGTTGGCGTA GGCATTGGT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadH-742_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTAGCTACGCAAACCACGTCC ACGTTGGAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadH-53_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGTCTGCACCCAGATACGCA TAAAGCGAT CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadH-88_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AAAGTAAAGGGTCATGGTGAT GACTGGCGG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadH-463_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGCAAACACACCGTTCAGCAA ACCCGCAAG CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadH-519_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TACGTGAGTGGCGTTAACACA AAGGTTGGC CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-73

Exp-1_yadH-151_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGCATATAGCTGAAGCCATG CATATCGCC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadH-599_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TATAAACGATTGGGTTCAGGT GCGACAGCC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadH-297_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TATCCGGCAATAATGACGTGA GTCGGAACC CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadH-419_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGTGAGCACCAGCGTTAAGG CAACGAATA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadH-653_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGACCAGCGGAACATCATTGA TACCGAGGA CATTGC
GCTTCTCCGGAAACATTAGC

Exp-1_yadH-688_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCCACAATAAAGACCACCAG TACGCCAAA CATTGC
GCTTCTCCGGAAACATTAGC

FIG. 4-74

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 15/42559

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12P 19/34; C12Q 1/68; C07H 21/04 (2015.01)
CPC - C12Q 1/686; C12N 15/10; C12Q 1/6844; C07H 21/00
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8) - C12P 19/34; C12Q 1/68; C07H 21/04 (2015.01); USPC - 435/91 .2; 435/91 .1; 435/6. 1; 435/6.12; 536/23.1
CPC - C12Q 1/686; C12N 15/10; C12Q 1/6844; C07H 21/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
IPC(8) - C12P 19/34; C12Q 1/68; C07H 21/04 (2015.01); USPC - 435/91 .2; 435/91 .1; 435/6.1 ; 435/6.12; 536/23.1
CPC - C12Q 1/686; C12N 15/10; C12Q 1/6844; C07H 21/00 - see keyword below

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PubWEST(USPT.PGPB.EPAB.JPAB); PatBase; Medline, Google: oligonucleotide, amplification, multiplex, plurality, library, pool, unique, PCR, real-time PCR, RT-PCR, reverse transcribing, producing, RNA, DNA, cDNA, selective, degrading, alkaline hydrolysis, RNase H, RNase T, DNA polymerase, Taq, RNA polymerase, T7 polymerase, fragment, aptamer, length

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 2012/0035065 A1 (SMOLKE et al.) 09 February 2012 (09.02.2012), para [0007], [0032], [0047], [0061], [0083], [0091], [0104], [01 17], [01 19], [0121], [0122], [0123], [0124], [0125], and [0147]	1-2, 73-74 ---------- 3, 75
Y	US 201 3/0096014 A1 (ANDERSEN et al.) 18 April 2013 (18.04.2013), para [0005], [0006], [0010], [0035], [0061], and [0067]	3, 75
A	FAKRUDDIN et al., NUCLEIC ACID SEQUENCE BASED AMPLIFICATION (NASBA)-PROSPECTS AND APPLICATIONS. Intern J Life Sci Pharm Res. 2012, Vol 2(1), p. L106-121 . Entire documentation, especially Abstract; and p. L108, Fig 1	1-3, 73-75
A	SHIELS et al., RNA-DNA Hybrids Containing Damaged DNA are Substrates for RNase H. Bioorg Med Chem Lett. 2001 , Vol. 11(19), p. 2623-6. Entire documentation, especially Abstract; and pg 2625, col 1, top para	1-3, 73-75

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance, the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 September 2015 (10.09.2015)	0 9 OCT 2015
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/2 10 (second sheet) (January 201 5)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 15/42559

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 4-72, 76-142
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. **III** Observations where unity of invention is lacking (Continuation of item **3** of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (2)) (January 2015)

PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION CONCERNING TRANSMITTAL OF COPY OF INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY (CHAPTER I OF THE PATENT COOPERATION TREATY)
(PCT Rule 44bis.1(c))

To:

CHEN, Tani
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206
ETATS-UNIS D'AMERIQUE

Date of mailing *(day/month/year)* 23 May 2019 (23.05.2019)	
Applicant's or agent's file reference H0498-70630	**IMPORTANT NOTICE**

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US2017/060570	08 November 2017 (08.11.2017)	08 November 2016 (08.11.2016)

Applicant
PRESIDENT AND FELLOWS OF HARVARD COLLEGE

The International Bureau transmits herewith a copy of the international preliminary report on patentability (Chapter I of the Patent Cooperation Treaty).

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No. +41 22 338 82 70

Authorized officer
Simin Baharlou

e-mail: pct.team9@wipo.int

Form PCT/IB/326 (January 2004)

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY
(Chapter I of the Patent Cooperation Treaty)

(PCT Rule 44*bis*)

Applicant's or agent's file reference H0498-70630	**FOR FURTHER ACTION**	See item 4 below
International application No. PCT/US2017/060570	International filing date *(day/month/year)* 08 November 2017 (08.11.2017)	Priority date *(day/month/year)* 08 November 2016 (08.11.2016)
International Patent Classification (8th edition unless older edition indicated) See relevant information in Form PCT/ISA/237		
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE		

1. This international preliminary report on patentability (Chapter I) is issued by the International Bureau on behalf of the International Searching Authority under Rule 44 *bis*.1(a).

2. This REPORT consists of a total of 7 sheets, including this cover sheet.

   In the attached sheets, any reference to the written opinion of the International Searching Authority should be read as a reference to the international preliminary report on patentability (Chapter I) instead.

3. This report contains indications relating to the following items:

☒	Box No. I	Basis of the report
☐	Box No. II	Priority
☒	Box No. III	Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
☐	Box No. IV	Lack of unity of invention
☒	Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
☐	Box No. VI	Certain documents cited
☐	Box No. VII	Certain defects in the international application
☐	Box No. VIII	Certain observations on the international application

4. The International Bureau will communicate this report to designated Offices in accordance with Rules 44*bis*.3(c) and 93*bis*.1 but not, except where the applicant makes an express request under Article 23(2), before the expiration of 30 months from the priority date (Rule 44*bis* .2).

	Date of issuance of this report 14 May 2019 (14.05.2019)
The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland Facsimile No. +41 22 338 82 70	Authorized officer Simin Baharlou e-mail: pct.team9@wipo.int

Form PCT/IB/373 (January 2004)

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To: TANI CHEN
WOLF, GREENFIELD & SACKS, P.C.
600 ATLANTIC AVENUE
BOSTON, MA 02210-2206

PCT

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing *(day/month/year)*	05 FEB 2018

Applicant's or agent's file reference	FOR FURTHER ACTION
H0498-70630	See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US 17/60570	08 November 2017 (08.11.2017)	08 November 2016 (08.11.2016)

International Patent Classification (IPC) or both national classification and IPC
IPC(8) - G06F 19/12 (2018.01)
CPC - G06F 19/12

Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☒ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☐ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☐ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA/US	Date of completion of this opinion	Authorized officer
Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	10 January 2018	Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/237 (cover sheet) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 17/60570

Box No. I Basis of this opinion

1. With regard to the **language**, this opinion has been established on the basis of:
   - ☒ the international application in the language in which it was filed.
   - ☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a)).

3. ☒ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:
   - a. ☒ forming part of the international application as filed:
     - ☒ in the form of an Annex C/ST.25 text file.
     - ☐ on paper or in the form of an image file.
   - b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
   - c. ☐ furnished subsequent to the international filing date for the purposes of international search only:
     - ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).
     - ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 17/60570

Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☒ claims Nos. 5-64, 68-118, 122-150, 154-169

because:

☐ the said international application, or the said claims Nos. _____ relate to the following subject matter which does not require an international search *(specify)*:

☒ the description, claims or drawings *(indicate particular elements below)* or said claims Nos. 5-64,68-118,122-150,154-169 are so unclear that no meaningful opinion could be formed *(specify)*:

Claims 5-64, 68-118, 122-150, 154-169 are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

☐ the claims, or said claims Nos. _____ are so inadequately supported by the description that no meaningful opinion could be formed *(specify)*:

☒ no international search report has been established for said claims Nos. 5-64, 68-118, 122-150, 154-169

☐ a meaningful opinion could not be formed without the sequence listing; the applicant did not, within the prescribed time limit:

☐ furnish a sequence listing in the form of an Annex C/ST.25 text file, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.

☐ furnish a sequence listing on paper or in the form of an image file complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.

☐ pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rule 13*ter*.1(a) or (b).

☐ See Supplemental Box for further details.

Form PCT/ISA/237 (Box No. III) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 17/60570

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	4, 66, 67, 119-121	YES
	Claims	1-3, 65, 151-153	NO
Inventive step (IS)	Claims	None	YES
	Claims	1-4, 65-67, 119-121, 151-153	NO
Industrial applicability (IA)	Claims	1-4, 65-67, 119-121, 151-153	YES
	Claims	None	NO

2. Citations and explanations:

Claims 1-3 and 65 lack novelty under PCT Article 33(2) as being anticipated by US 2015/0087001 A1 to California Institute Of Technology (hereinafter 'Caltech').

Regarding Claim 1, Caltech teaches a method, comprising:
exposing a sample to a plurality of nucleic acid probes (para [0017] - '100 um-thick mouse brain slices were hybridized with twenty-four 20 mer oligonucleotide probes towards beta-actin mRNA labeled with Alexafluor 594.');
polymerizing a gel within the sample (claim 1 - 'applying a hydrogel monomer solution comprising acrylamide and phosphate buffered saline (PBS) to the fixed tissue, thereby forming hydrogel treated tissue.'; para [0025] - 'Immediately prior to hydrogel polymerization, the entire perfusion chamber (with catheter-connected subject) is sealed within a ziplock bag for tissue degassing');
anchoring a target to the gel; clearing non-targets from the sample (claim 1 - 'applying a hydrogel monomer solution comprising acrylamide and phosphate buffered saline (PBS) to the fixed tissue, thereby forming hydrogel treated tissue.'; para [0016] - 'FIG. 1 demonstrates, in accordance with an embodiment of the invention, PACT clearing of A4P0 tissue-hydrogel hybrid achieves optimal transparency and immunohistochemistry compatibility across organs.'; [0056] 'native proteins including those maintaining the structural integrity of tissue samples, retained by the tissue-hydrogel matrix during clearing'); and
determining the targets within the gel by determining binding of the nucleic acid probes by imaging (para [0017] - 'twenty-four 20 mer oligonucleotide probes towards beta-actin mRNA labeled with Alexafluor 594...An abundant number of diffraction limited spots corresponding to single beta-actin mRNAs (red) were readily detected up to 30 um in depth under 589 nm illumination....the images were processed from raw data'; [0018] - ' After 24-hour incubation in RIMS, PARS-CSF brain slices were sufficiently cleared for imaging without further sectioning....Images show native eGFP fluorescence in 500 um PARS-CSF cleared coronal brain slice').

Regarding Claim 2, Caltech teaches the method of claim 1, wherein the target is a nuclelc acid (para [0017] - 'FIG. 2 demonstrates, in accordance with an embodiment of the invention, detection of individual mRNA transcripts in PACT tissue sections by smFISH.').

Regarding Claim 3, Caltech teaches the method of any one of claims 1 or 2, wherein the target comprises RNA (para [0017] - 'FIG. 2 demonstrates, in accordance with an embodiment of the invention, detection of individual mRNA transcripts in PACT tissue sections by smFISH.').

Regarding Claim 65, Caltech teaches a method, comprising:
exposing a sample to a plurality of nucleic acid probes (para [0017] - '100 um-thick mouse brain slices were hybridized with twenty-four 20 mer oligonucleotide probes towards beta-actin mRNA labeled with Alexafluor 594.');
polymerizing a gel within the sample (claim 1 - 'applying a hydrogel monomer solution comprising acrylamide and phosphate buffered saline (PBS) to the fixed tissue, thereby forming hydrogel treated tissue.'; para [0025] - 'Immediately prior to hydrogel polymerization, the entire perfusion chamber (with catheter-connected subject) is sealed within a ziplock bag for tissue degassing');
anchoring a target to the gel (claim 1 - 'applying a hydrogel monomer solution comprising acrylamide and phosphate buffered saline (PBS) to the fixed tissue, thereby forming hydrogel treated tissue.'; para [0016] - 'FIG. 1 demonstrates, in accordance with an embodiment of the invention, PACT clearing of A4P0 tissue-hydrogel hybrid achieves optimal transparency and immunohistochemistry compatibility across organs.'; [0056] 'native proteins including those maintaining the structural integrity of tissue samples, retained by the tissue-hydrogel matrix during clearing');
reducing background fluorescence within the sample (para [0017] - '(B) Compared to PACT cleared slices, smFISH in uncleared brain slices showed significantly decreased contrast....Comparison of background intensity between uncleared and PACT-cleared tissue illustrates the significant reduction of background fluorescence in PACT-cleared tissue'); and
imaging the nucleic acid probes (para [0017] - 'twenty-four 20 mer oligonucleotide probes towards beta-actin mRNA labeled with Alexafluor 594...An abundant number of diffraction limited spots corresponding to single beta-actin mRNAs (red) were readily detected up to 30 um in depth under 589 nm illumination....the images were processed from raw data'; [0018] - ' After 24-hour incubation in RIMS, PARS-CSF brain slices were sufficiently cleared for imaging without further sectioning....Images show native eGFP fluorescence in 500 um PARS-CSF cleared coronal brain slice').

---Please see continuation in first supplemental box----------

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 17/60570

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:
Box V2 Citations and Explanations

Claims 151-153 lack novelty under PCT Article 33(2) as being anticipated by US 2016/0304952 A1 to Massachusetts Institue Of Technology et al., (hereinafter 'MIT').

Regarding Claim 151, MIT teaches a method, comprising:
embedding at least a portion of a sample within a matrix (Claim 1 - 'b) embedding the biological sample comprising the target nucleic acids and attached small molecule linker or nucleic acid adaptor in a swellable material wherein the small molecule linker or the nucleic acid adaptor is linked to the target nucleic acids present in the sample and to the swellable material'; para [0037] - 'Monomers or oligomers can comprise one or more substituted or unsubstituted methacrylates, acrylates, acrylamides, methacrylamides,');
immobilizing targets to the matrix (Claim 1 - 'a) linking target nucleic acids present in the biological sample with a small molecule linker or a nucleic acid adaptor capable of linking to a target nucleic acid and to a swellable material'; para [0029] - 'Additionally, acrylamide modified oligonucleotide primers may be covalently fixed within a swellable material such as a polyacrylate gel');
clearing non-targets from the matrix (Claim 1 - 'c) digesting proteins present in the biological sample'; para [0047] 'The degree of digestion can be sufficient to compromise the integrity of the mechanical structure of the sample or it can be complete to the extent that the sample-swellable material complex is rendered substantially free of the sample. Preferably the disruption of the physical structure of the sample is protein digestion of the proteins contained in the biological sample'); and imaging the targets within the matrix (para [0050] - 'The swollen material with the embedded biological sample can be imaged on any optical microscope'; para [0011] - 'FIG. 3 shows microscopy images showing in situ sequencing').

Regarding Claim 152, MIT teaches the method of claim 151, wherein the matrix comprises a polymer (para [0037] - 'Monomers or oligomers can comprise one or more substituted or unsubstituted methacrylates, acrylates, acrylamides, methacrylamides,').

Regarding Claim 153, MIT teaches the method of any one of claims 151 or 152, wherein the matrix comprises a gel (para [0010] - 'Panel B depicts that biological specimens are embedded in a swellable gel material').

Claims 4, 66-67 lack an inventive step under PCT Article 33(3) as being obvious over Caltech.

Regarding Claim 4, Caltech teaches the method of any one of claims 1 or 2, yet does not expressly teach the target comprises DNA. Since Caltech teaches the method is intended for detection of DNA among other analytes (para [0037] - 'Thus motivation to improve tissue clearing protocols is sustained around three main objectives: 1) efficient clearing of both central organs and peripheral tissues; 2) preservation of cellular and subcellular structures of multiple organ types; and 3) compatibility with endogenous fluorescent protein expression and post-hoc detection of DNA, RNA, and proteins.'), it would have been obvious to one of ordinary skill in the art that the method of Caltech could be used to detect target DNA.

Regarding Claim 66, Caltech teaches the method of claim 65, yet does not expressly teach wherein during imaging, the gel has not expanded by more than 3x. Caltech does teach limited hydrogel swelling (para [0063] - 'Indeed, PARS-based clearing of rodent brains was accomplished with limited hydrogel swelling and tissue expansion during clearing (FIG. 4C-4D, 11A)'). Since Caltech teaches tissue-hydrogel matrices exhibited less than 300% (i.e., less than 3x) tissue weight and volume changes during imaging (para [0054] - 'Indeed, tissue-hydrogel matrices that were prepared using acrylamide alone (A4P0) exhibited tissue weight and volume changes of 174% and 223%, respectively (FIG. 1E) over A4P4 counterparts'), it would have been obvious to one of ordinary skill in the art that the gel had not expanded by more than 3x since the overall tissue-hydrogel matrices exhibited less than 3x expansion.

Regarding Claim 67, Caltech teaches the method of any one of claims 65 or 66, yet does not expressly teach wherein during imaging, the gel has not expanded by more than 1.5x. Caltech does teach limited hydrogel swelling (para [0063] - 'Indeed, PARS-based clearing of rodent brains was accomplished with limited hydrogel swelling and tissue expansion during clearing (FIG. 4C-4D, 11A)'). Since Caltech teaches tissue-hydrogel matrices exhibited a variable range of swelling based on their reagent composition (para [0054] - 'Indeed, tissue-hydrogel matrices that were prepared using acrylamide alone (A4P0) exhibited tissue weight and volume changes of 174% and 223%, respectively (FIG. 1E) over A4P4 counterparts'), it would have been obvious to one of ordinary skill in the art that by varying the reagent composition such as the percentage of paraformaldehyde vs acrylamide, one could identify by routine experimentation a reagent composition wherein the gel expansion does not exceed 1.5x.

----Please see continuation in next supplemental box----------

Form PCT/ISA/237 (Supplemental Box) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 17/60570

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:
Box V2 Citations and Explanations

Claims 119-121 lack an inventive step under PCT Article 33(3) as being obvious over MIT in view of the article entitled 'High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization' by Moffitt et al., (hereinafter 'Moffitt').

Regarding Claim 119, MIT teaches a method (Claim 1 - 'A method for in-situ sequencing of target nucleic acids present in a biological sample'), comprising:
exposing a sample to a plurality of anchor nucleic acid probes (Claim 1 - 'a) linking target nucleic acids present in the biological sample with a small molecule linker or a nucleic acid adaptor capable of linking to a target nucleic acid and to a swellable material'; para [0029] - ' For example, anchorable DNA adaptors may be ligated to the 3' ends of the RNAs in the sample with RNA ligases, such as T4 RNA ligase, or may be attached via a chemical linker such as a reactive amine group capable of reacting with target nucleic acid.');
embedding at least a portion of the sample within a polyacrylamide gel (Claim 1 - 'b) embedding the biological sample comprising the target nucleic acids and attached small molecule linker or nucleic acid adaptor in a swellable material wherein the small molecule linker or the nucleic acid adaptor is linked to the target nucleic acids present in the sample and to the swellable material'; para [0037] - 'Monomers or oligomers can comprise one or more substituted or unsubstituted methacrylates, acrylates, acrylamides, methacrylamides,')
immobilizing at least some of the anchor nucleic acid probes to the polyacrylamide gel (para [0029] - 'Additionally, acrylamide modified oligonucleotide primers may be covalently fixed within a swellable material such as a polyacrylate gel'); clearing proteins and/or lipids and/or DNA and/or extracellular matrix and/or RNA molecules from the sample (Claim 1 - 'c) digesting proteins present in the biological sample'; para [0047] 'The degree of digestion can be sufficient to compromise the integrity of the mechanical structure of the sample or it can be complete to the extent that the sample-swellable material complex is rendered substantially free of the sample. Preferably the disruption of the physical structure of the sample is protein digestion of the proteins contained in the biological sample'); and determining sequence of target mRNA by imaging the polyacrylamide gel (para [0050] - 'The swollen material with the embedded biological sample can be imaged on any optical microscope'; para [0011] - 'FIG. 3 shows microscopy images showing in situ sequencing'). MIT does not expressly teach exposing a sample to a plurality of MERFISH nucleic acid probes; or determining binding of the MERFISH nucleic acid probes. Moffitt teaches exposing a sample to a plurality of MERFISH nucleic acid probes and determining binding of the MERFISH nucleic acid probes (Abstract - '. Here we report advances that lead to a drastic increase in the measurement throughput of multiplexed error-robust fluorescence in situ hybridization (MERFISH), an image based approach to single-cell transcriptomics. In MERFISH, RNAs are identified via a combinatorial labeling approach that encodes RNA species with error-robust barcodes followed by sequential rounds of single-molecule fluorescence in situ hybridization (smFISH) to read out these barcodes.'). Since both MIT and Moffitt teach methods based on smFISH, it would have been obvious to one of ordinary skill in the art that the MERFISH probes of Moffitt could be used with the method of MIT to image target mRNAs in situ.

Regarding Claim 120, MIT in view of Moffitt teaches the method of claim 119, wherein MIT further teaches the polyacrylamide gel comprises anchor probes incorporated within the polyacrylamide gel (para [0029] - 'Additionally, acrylamide modified oligonucleotide primers may be covalently fixed within a swellable material such as a polyacrylate gel').

Regarding Claim 121, MIT in view of Moffitt teaches the method of any one of claims 119 or 120, wherein MIT further teaches clearing proteins and/or lipids from the sample comprises removing proteins and/or lipids from the sample (Claim 1 - 'c) digesting proteins present in the biological sample').

Claims 1-4, 65-67, 119-121 and 151-153 have industrial applicability as defined by PCT Article 33(4) because the subject matter can be made or used in industry.

PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION CONCERNING TRANSMITTAL OF COPY OF INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY (CHAPTER I OF THE PATENT COOPERATION TREATY)
(PCT Rule 44bis.1(c))

To:

CHEN, Tani
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206
ETATS-UNIS D'AMERIQUE

Date of mailing *(day/month/year)* 23 May 2019 (23.05.2019)		
Applicant's or agent's file reference H0498.70645	**IMPORTANT NOTICE**	
International application No. PCT/US2017/060558	International filing date *(day/month/year)* 08 November 2017 (08.11.2017)	Priority date *(day/month/year)* 08 November 2016 (08.11.2016)
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE		

The International Bureau transmits herewith a copy of the international preliminary report on patentability (Chapter I of the Patent Cooperation Treaty).

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No. +41 22 338 82 70

Authorized officer
Mineko Mohri

e-mail: pct.team8@wipo.int

Form PCT/IB/326 (January 2004)

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY
(Chapter I of the Patent Cooperation Treaty)

(PCT Rule 44*bis*)

Applicant's or agent's file reference H0498.70645	**FOR FURTHER ACTION**	See item 4 below
International application No. PCT/US2017/060558	International filing date *(day/month/year)* 08 November 2017 (08.11.2017)	Priority date *(day/month/year)* 08 November 2016 (08.11.2016)
International Patent Classification (8th edition unless older edition indicated) See relevant information in Form PCT/ISA/237		
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE		

1. This international preliminary report on patentability (Chapter I) is issued by the International Bureau on behalf of the International Searching Authority under Rule 44 *bis*.1(a).

2. This REPORT consists of a total of 6 sheets, including this cover sheet.

   In the attached sheets, any reference to the written opinion of the International Searching Authority should be read as a reference to the international preliminary report on patentability (Chapter I) instead.

3. This report contains indications relating to the following items:

☒	Box No. I	Basis of the report
☐	Box No. II	Priority
☒	Box No. III	Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
☐	Box No. IV	Lack of unity of invention
☒	Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
☐	Box No. VI	Certain documents cited
☐	Box No. VII	Certain defects in the international application
☐	Box No. VIII	Certain observations on the international application

4. The International Bureau will communicate this report to designated Offices in accordance with Rules 44*bis*.3(c) and 93*bis*.1 but not, except where the applicant makes an express request under Article 23(2), before the expiration of 30 months from the priority date (Rule 44*bis* .2).

	Date of issuance of this report 14 May 2019 (14.05.2019)
The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland Facsimile No. +41 22 338 82 70	Authorized officer Mineko Mohri e-mail: pct.team8@wipo.int

Form PCT/IB/373 (January 2004)

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To: TANI CHEN
WOLF, GREENFIELD & SACKS, P.C.
600 ATLANTIC AVENUE
BOSTON, MA 02210-2206

PCT

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing *(day/month/year)*	05 FEB 2018
Applicant's or agent's file reference H0498.70645	**FOR FURTHER ACTION** See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US 17/60558	08 November 2017 (08.11.2017)	08 November 2016 (08.11.2016)

International Patent Classification (IPC) or both national classification and IPC
IPC(8) - G01N 1/36, C12Q 1/68 (2017.01)
CPC - C12Q 1/6841, C12Q 1/6806, G01N 2001/364, G01N 1/36

Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☒ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☐ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☐ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA/US	Date of completion of this opinion	Authorized officer
Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	04 January 2018	Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/237 (cover sheet) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 17/60558

Box No. I Basis of this opinion

1. With regard to the **language**, this opinion has been established on the basis of:

   ☒ the international application in the language in which it was filed.

   ☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a)).

3. ☐ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:

   a. ☐ forming part of the international application as filed:

      ☐ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 17/60558

Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☒ claims Nos. 5-57

because:

☐ the said international application, or the said claims Nos. _____ relate to the following subject matter which does not require an international search *(specify)*:

☒ the description, claims or drawings *(indicate particular elements below)* or said claims Nos. 5-57 are so unclear that no meaningful opinion could be formed *(specify)*:

because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

☐ the claims, or said claims Nos. _____ are so inadequately supported by the description that no meaningful opinion could be formed *(specify)*:

☒ no international search report has been established for said claims Nos. 5-57

☐ a meaningful opinion could not be formed without the sequence listing; the applicant did not, within the prescribed time limit:

- ☐ furnish a sequence listing in the form of an Annex C/ST.25 text file, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.
- ☐ furnish a sequence listing on paper or in the form of an image file complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.
- ☐ pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rule 13*ter*.1(a) or (b).

☐ See Supplemental Box for further details.

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 17/60558

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	59	YES
	Claims	1-4, 58, 60-61	NO
Inventive step (IS)	Claims	None	YES
	Claims	1-4, 58-61	NO
Industrial applicability (IA)	Claims	1-4, 58-61	YES
	Claims	None	NO

2. Citations and explanations:

Claims 1-4, 58, 60-61 lack novelty under PCT Article 33(2) as being anticipated by US 2016/0304952 A1 to Massachusetts Institute Of Technology et al. (hereafter, MIT).

Regarding claim 1, MIT teaches a method, comprising: embedding cells within an expandable material (para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: ... b) embedding the biological sample in a swellable material wherein the small molecule linker or nucleic acid adaptor is attached both to the target nucleic acids present in the sample and to the swellable material; para [0027]; As used herein, the term "biological sample" means any biological sample that comprises, or is believed to comprise, nucleic acid sequences including, but not limited to, cDNA, mRNA and genomic DNA. Such samples may include tissues containing a variety of cells, a single cell or organelles. Suitable biological samples include, but are not limited to, a blood sample, biopsy specimen, tissue explant, organ culture, biological fluid or any other tissue or cell preparation, or fraction or derivative thereof or isolated therefrom); immobilizing nucleic acids from the cells to the expandable material (para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: a) attaching target nucleic acids present in the biological sample with a small molecule linker or nucleic acid adapter; b) embedding the biological sample in a swellable material wherein the small molecule linker or nucleic acid adaptor is attached both to the target nucleic acids present in the sample and to the swellable material); expanding the expandable material (para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: ... d) swelling the swellable material to form a first enlarged biological sample that is enlarged as compared to the biological sample); exposing the expandable material to a plurality of nucleic acid probes (para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: ... (g) sequencing the nucleic acids present in the first enlarged sample; para [0032]; General sequencing methods known in the art, such as sequencing by ... fluorescent in situ sequencing (FISSEQ); para [0033]; FISSEQ is a method whereby DNA is extended by adding a single type of fluorescently-labelled nucleotide triphosphate to the reaction); and determining binding of the nucleic acid probes to the immobilized nucleic acids (para [0033]; FISSEQ is a method whereby DNA is extended by adding a single type of fluorescently-labelled nucleotide triphosphate to the reaction; washing away unincorporated nucleotide, detecting incorporation of the nucleotide by measuring fluorescence, and repeating the cycle).

Regarding claim 2, MIT teaches the method of claim 1, wherein immobilizing the nucleic acids from the cells to the expandable material comprises immobilizing the nucleic acids at a single point to the expandable material (para [0029]; anchorable DNA adaptors may be ligated to the 3' ends of the RNAs in the sample with RNA ligases; (since the RNA samples are ligated at the 3' ends, it would follow that the RNA samples are only immobilized at one point which is the 3' end)).

Regarding claim 3, MIT teaches the method of claim 1, comprising immobilizing the nucleic acids to the expandable material (para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: a) attaching target nucleic acids present in the biological sample with a small molecule linker or nucleic acid adapter; b) embedding the biological sample in a swellable material wherein the small molecule linker or nucleic acid adaptor is attached both to the target nucleic acids present in the sample and to the swellable material).

Regarding claim 4, MIT teaches the method of claim 3, comprising immobilizing the nucleic acids at a 5' end of the nucleic acids to the expandable material (para [0029]; acrylamide modified oligonucleotide primers may be covalently fixed within a swellable material such as a polyacrylate gel. As used herein, the term "acrylamide modified" in reference to an oligonucleotide means that the oligonucleotide has an acrylamide moiety attached to the 5' end of the molecule).

--- V.2 Citations and explanations ---

Form PCT/ISA/237 (Box No. V) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 17/60558

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**
Continuation of:
--- Continued in next Supplemental Box ---

Regarding claim 58, MIT teaches a method, comprising: immobilizing a plurality of targets to an expandable material (para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: a) attaching target nucleic acids present in the biological sample with a small molecule linker or nucleic acid adapter; b) embedding the biological sample in a swellable material wherein the small molecule linker or nucleic acid adaptor is attached both to the target nucleic acids present in the sample and to the swellable material), wherein at least 50% of the plurality of targets immobilized to the expandable material are immobilized at single points (para [0029]; anchorable DNA adaptors may be ligated to the 3' ends of the RNAs in the sample with RNA ligases; (since the RNA samples are ligated at the 3' ends, it would follow that all the RNA samples are only immobilized at a single point which is the 3' end)); and expanding the expandable material (para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: ... d) swelling the swellable material to form a first enlarged biological sample that is enlarged as compared to the biological sample).

Regarding claim 60, MIT teaches an article, comprising: an expandable material (para [0037]; Preferably, the swellable material is a swellable polymer or hydrogel), comprising an embedded cell and a plurality of nucleic acids immobilized to the expandable material (para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: a) attaching target nucleic acids present in the biological sample with a small molecule linker or nucleic acid adapter; b) embedding the biological sample in a swellable material wherein the small molecule linker or nucleic acid adaptor is attached both to the target nucleic acids present in the sample and to the swellable material), wherein at least 50% of the plurality of nucleic acids immobilized to the expandable material are immobilized at single points (para [0029]; anchorable DNA adaptors may be ligated to the 3' ends of the RNAs in the sample with RNA ligases; (since the RNA samples are ligated at the 3' ends, it would follow that all the RNA samples are only immobilized at a single point which is the 3' end)).

Regarding claim 61, MIT teaches an article, comprising: a polymer comprising an expanded cell and a plurality of nucleic acids immobilized to the polymer (para [0037]; Preferably, the swellable material is a swellable polymer or hydrogel; para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: a) attaching target nucleic acids present in the biological sample with a small molecule linker or nucleic acid adapter; b) embedding the biological sample in a swellable material wherein the small molecule linker or nucleic acid adaptor is attached both to the target nucleic acids present in the sample and to the swellable material ... d) swelling the swellable material to form a first enlarged biological sample that is enlarged as compared to the biological sample), wherein the cell is expanded to at least 5 times its normal size within the polymer (para [0049]; In one embodiment, the addition of water allows for the embedded sample to expand at least 3 times, preferably 4 times, preferably 5 times, or more its original size in three-dimensions), and wherein at least 50% of the plurality of nucleic acids immobilized to the expandable material are immobilized at single points (para [0029]; anchorable DNA adaptors may be ligated to the 3' ends of the RNAs in the sample with RNA ligases; (since the RNA samples are ligated at the 3' ends, it would follow that all the RNA samples are only immobilized at a single point which is the 3' end)).

Claim 59 lacks an inventive step under PCT Article 33(3) as being obvious over MIT.

Regarding claim 59, MIT teaches a method, comprising: immobilizing a plurality of nucleic acids to an expandable material (para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: a) attaching target nucleic acids present in the biological sample with a small molecule linker or nucleic acid adapter; b) embedding the biological sample in a swellable material wherein the small molecule linker or nucleic acid adaptor is attached both to the target nucleic acids present in the sample and to the swellable material), wherein at least 50% of the plurality of nucleic acids immobilized to the expandable material are immobilized at single points (para [0029]; anchorable DNA adaptors may be ligated to the 3' ends of the RNAs in the sample with RNA ligases; (since the RNA samples are ligated at the 3' ends, it would follow that all the RNA samples are only immobilized at a single point which is the 3' end)); exposing the expandable material to a plurality of nucleic acid probes (para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: ... (g) sequencing the nucleic acids present in the first enlarged sample; para [0032]; General sequencing methods known in the art, such as sequencing by ... fluorescent in situ sequencing (FISSEQ); para [0033]; FISSEQ is a method whereby DNA is extended by adding a single type of fluorescently-labelled nucleotide triphosphate to the reaction); expanding the expandable material (para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: ... d) swelling the swellable material to form a first enlarged biological sample that is enlarged as compared to the biological sample); and determining binding of the nucleic acid probes to the immobilized nucleic acids (para [0033]; FISSEQ is a method whereby DNA is extended by adding a single type of fluorescently-labelled nucleotide triphosphate to the reaction; washing away unincorporated nucleotide, detecting incorporation of the nucleotide by measuring fluorescence, and repeating the cycle);
but does not specifically teach wherein the step of exposing the expandable material to a plurality of nucleic acid probes occurs before expanding the expandable material. Nonetheless, since MIT teaches the steps, it would have been obvious to one of ordinary skill in the art to trivially switch the order of labelling steps and expanding the expandable material during routine experimentation to optimize the method.

Claims 1-4, 58-61 have industrial applicability as defined by PCT Article 33(4) because the subject matter can be made or used in industry.

Electronic Patent Application Fee Transmittal

Application Number:	15329683
Filing Date:	27-Jan-2017
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Filer:	Tani Chen/Catherine Borgnis
Attorney Docket Number:	H0498.70509US02

Filed as Small Entity

Filing Fees for U.S. National Stage under 35 USC 371

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
SUBMISSION- INFORMATION DISCLOSURE STMT	2806	1	120	120
		Total in USD ($)		**120**

Electronic Acknowledgement Receipt

EFS ID:	36717148
Application Number:	15329683
International Application Number:	
Confirmation Number:	7428
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Customer Number:	86110
Filer:	Tani Chen/Catherine Borgnis
Filer Authorized By:	Tani Chen
Attorney Docket Number:	H0498.70509US02
Receipt Date:	29-JUL-2019
Filing Date:	27-JAN-2017
Time Stamp:	15:52:28
Application Type:	U.S. National Stage under 35 USC 371

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$120
RAM confirmation Number	E20197SF53189790
Deposit Account	232825
Authorized User	Catherine Borgnis
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows: 37 CFR 1.17 (Patent application and reexamination processing fees)	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	H049870509US02-TRN-TC.pdf	26004 df7b76ceb1b0767f283253d47b09a630034603a9	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	H049870509US02-FEE-TC.pdf	32725 85bfec60175297b1daf8104e734d6c8cc10ea79c	no	1
Warnings:					
Information:					
3		H049870509US02-IDS-TC.pdf	37250 be943b8aea92627f17a5f2719e1024a1fbf3a3b5	yes	5

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Transmittal Letter	1	4
	Information Disclosure Statement (IDS) Form (SB08)	5	5

Warnings:					
Information:					
4	Foreign Reference	WO2008091296A2.pdf	908888 4041d1a930d24c45e83644c2da9345b46e1f3b7b	no	81
Warnings:					
Information:					
5	Foreign Reference	WO2016018963A9.pdf	4165958 cc76f48fd3074136094dd69c8f3fdd1e481af708	no	131
Warnings:					
Information:					

6	Non Patent Literature	LEVSKY_et_al_Science_2002_v297_p836.pdf	411229 8ab7ba3ddb1471a29081051a6dccc68301e414d0	no	5
Warnings:					
Information:					
7	Non Patent Literature	Player_et_al_J_Hist_Cyto_2001_v49_p603.pdf	538404 8b90714f13744dad66bb4822f90f7f418336cab0	no	9
Warnings:					
Information:					
8	Non Patent Literature	Zhen_et_al_Prenatal_Diag_1998_v18_p1181.pdf	136302 7d04d11e3794af1ff890387e3a34fd07ada25fe4	no	5
Warnings:					
Information:					
9	Other Reference-Patent/App/Search documents	H049870509CN00_CN_OA_mailed_01-16-2019_QC.pdf	7437403 bbe60157852751 88f5a93ca4f5763810356aba32	no	14
Warnings:					
Information:					
10	Other Reference-Patent/App/Search documents	H049870630WO00_IPRP_mailed_05-23-2019.PDF	409380 02334ffc214846d6b30238dc3d1982d4bc4334d8	no	8
Warnings:					
Information:					
11	Other Reference-Patent/App/Search documents	H049870645WO00_IPRP_mailed_05-23-2019.PDF	364518 e41f6002e517da35189c6a19065000969ea0b336	no	7
Warnings:					
Information:					
12	Fee Worksheet (SB06)	fee-info.pdf	30672 46f175c691a4070d1ff1150925d20293e1c456f6	no	2
Warnings:					
Information:					
			Total Files Size (in bytes):	14498733	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

FEE TRANSMITTAL

Complete if known	
Application Number	15/329,683-Conf. #7428
Filing Date	January 27, 2017
First Named Inventor	Xiaowei Zhuang
Examiner Name	F. W. Lu
Art Unit	1634
Practitioner Docket No.	H0498.70509US02

[X] Applicant asserts small entity status. See 37 CFR 1.27.

[ ] Applicant certifies micro entity status. See 37 CFR 1.29. Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

TOTAL AMOUNT OF PAYMENT ($)120.00

METHOD OF PAYMENT (check all that apply)

[ ] Check [X] Credit Card [ ] Money Order [ ] None [ ] Other (please identify): _____

[X] Deposit Account Deposit Account Number: 23/2825 Deposit Account Name: Wolf, Greenfield & Sacks, P.C.

For the above-identified deposit account, the Director is hereby authorized to (check all that apply):

[X] Charge fee(s) indicated below

[ ] Charge fee(s) indicated below, **except for the filing fee**

[X] Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17, **except for any excess claims fees or multiple dependent claim fee**

[ ] Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17

[X] Credit any overpayment of fee(s)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES (U = undiscounted fee; S = small entity fee; M = micro entity fee)

	FILING FEES			SEARCH FEES			EXAMINATION FEES			
Application Type	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**Fees Paid ($)**
Utility	300	150*	75	660	330	165	760	380	190	
Design	200	100	50	160	80	40	600	300	150	
Plant	200	100	50	420	210	105	620	310	155	
Reissue	300	150	75	660	330	165	2,200	1,100	550	
Provisional	280	140	70	0	0	0	0	0	0	

* The $150 small entity status filing fee for a utility application is further reduced to $75 for a small entity status applicant who files the application via EFS-Web.

2. EXCESS CLAIM FEES

Fee Description	**Undiscounted Fee ($)**	**Small Entity Fee ($)**	**Micro Entity Fee ($)**
Each claim over 20 (including Reissues)	100	50	25
Each independent claim over 3 (including Reissues)	460	230	115
Multiple dependent claims	820	410	205

Total Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**
_____ - 20 or HP =	_____ x	_____ =	_____

HP = highest number of total claims paid for, if greater than 20.

Multiple Dependent Claims	
Fee ($)	**Fee Paid ($)**
_____	_____

Indep. Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**
_____ - 3 or HP =	_____ x	_____ =	_____

HP = highest number of independent claims paid for, if greater than 3.

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper (excluding electronically filed sequence or computer listings under 37 CFR 1.52(e)), the application size fee due is $400 ($200 for small entity) ($100 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets	**Extra Sheets**	**Number of each additional 50 or fraction thereof**	**Fee ($)**	**Fee Paid ($)**
_____ - 100 =	_____ /50 =	_____ (round **up** to a whole number) x	_____ =	_____

4. OTHER FEE(S)

	Fees Paid ($)
Non-English specification, $130 fee (no small or micro entity discount)	
Non-electronic filing fee under 37 CFR 1.16(t) for a utility application, $400 fee ($200 small or micro entity)	
Other (e.g., late filing surcharge): 2806 Submission of an Information Disclosure Statement	120.00

SUBMITTED BY

Signature	/Tani Chen/	Registration No. (Attorney/Agent)	52,728	Telephone	617.646.8000
Name (Print/Type)	Tani Chen, Sc.D.			Date	July 29, 2019

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: ____July 29, 2019___ Electronic Signature for Catherine E. Borgnis: __/Catherine E. Borgnis/___

7343414.1

DOCKET NO.: H0498.70509US02

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	Xiaowei Zhuang
Application No.:	15/329,683
Confirmation No.:	7428
Filed:	January 27, 2017
For:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
Examiner:	Frank Wei Min Lu
Art Unit:	1634

CERTIFICATE OF ELECTRONIC FILING UNDER 37 C.F.R. § 1.8

The undersigned hereby certifies that this paper, along with any paper referred to as being attached or enclosed, is being transmitted via the Office electronic filing system in accordance with § 1.6(a)(4), on the 29th day of July, 2019.

_____/Catherine E. Borgnis/ _____
Catherine E. Borgnis

MAIL STOP AMENDMENT
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

STATEMENT FILED PURSUANT TO THE DUTY OF DISCLOSURE UNDER 37 C.F.R. §§ 1.56, 1.97 AND 1.98

Sir:

Pursuant to the duty of disclosure under 37 C.F.R. §§ 1.56, 1.97 and 1.98, the undersigned requests consideration of this Information Disclosure Statement.

PART I: Compliance with 37 C.F.R. § 1.97

This Information Disclosure Statement has been filed more than three months after the filing date of this application and after the mailing date of a first Office Action, but before the mailing date of any of a Final Action under 37 C.F.R. § 1.113, a Notice of Allowance under 37 C.F.R. § 1.311, or an action that otherwise closes prosecution in this application.

Please charge our Credit Card in the amount of $120.00 covering the fee set forth in 37 C.F.R. § 1.17(p).

PART II: Information Cited

The undersigned hereby makes of record in the above-identified application the information listed on the attached form PTO-1449 (modified PTO/SB/08). The order of presentation of the references should not be construed as an indication of the importance of the references.

The undersigned hereby makes the following additional information of record in the above-identified application.

The undersigned would like to bring to the Examiner's attention the following co-pending applications that may contain subject matter related to this application:

Serial No.	Filing Date	Inventor(s)	Docket No.
*16/253,919	1/22/2019	Zhuang et al.	H0498.70510US02
*16/347,874	5/7/2019	Zhuang et al.	H0498.70630US01
*16/348,071	5/7/2019	Zhuang et al.	H0498.70645US01

*A copy of this reference is not provided as the Office has waived the requirement under 37 C.F.R. § 1.98(a)(2)(iii) for submitting a copy of a cited U.S. patent application if it is scanned to the Image File Wrapper system and is available on Private PAIR.

The undersigned would like to bring to the Examiner's attention the enclosed search report or other communication from a corresponding or related International or Foreign National Application:

Serial No.	Date of Mailing	Type(s) of Communication	Docket No.
CN 201580052678.7	**01-16-2019**	**Chinese Office Action**	**H0498.70509CN00**
PCT/US17/60570	**05-23-2019**	**International Preliminary Report on Patentability**	**H0498.70630WO00**
PCT/US17/60558	**05-23-2019**	**International Preliminary Report on Patentability**	**H0498.70645WO00**

PART III: Remarks

Documents cited anywhere in the Information Disclosure Statement are enclosed unless otherwise indicated. It is respectfully requested that:

1. The Examiner consider completely the cited information, along with any other information, in reaching a determination concerning the patentability of the present claims;

2. The enclosed form PTO-1449 (modified PTO/SB/08) be signed by the Examiner to evidence that the cited information has been fully considered by the United States Patent and Trademark Office during the examination of this application;

3. The citations for the information be printed on any patent which issues from this application.

By submitting this Information Disclosure Statement, the undersigned makes no representation that a search has been performed, of the extent of any search performed, or that more relevant information does not exist.

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, material to patentability as defined in 37 C.F.R. § 1.56(b).

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, in fact, prior art as defined by 35 U.S.C. § 102.

Notwithstanding any statements by the undersigned, the Examiner is urged to form his or her own conclusion regarding the relevance of the cited information.

An early and favorable action is hereby requested.

The Director is hereby authorized to charge any deficiency or credit any overpayment in the fees occasioned by the filing of this Information Disclosure Statement to our Deposit Account No. 23/2825 under Docket No. H0498.70509US02 from which the undersigned is authorized to draw.

Respectfully submitted,

By: /Tani Chen/

Tani Chen, Sc.D., Reg. No. 52,728
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
Telephone: (617) 646-8000

Docket No.: H0498.70509US02
Date: July 29, 2019
xNDDx

PTO/SB/06 (09-11)
Approved for use through 1/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PATENT APPLICATION FEE DETERMINATION RECORD Substitute for Form PTO-875	Application or Docket Number 15/329,683	Filing Date 01/27/2017	☐ To be Mailed

ENTITY: ☐ LARGE ☑ SMALL ☐ MICRO

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA		RATE ($)	FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A		N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A		N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 =	*		x $40 =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 =	*		x $210 =	
☐ APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).				
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))					
* If the difference in column 1 is less than zero, enter "0" in column 2.				TOTAL	

APPLICATION AS AMENDED - PART II

		(Column 1)		(Column 2)	(Column 3)			
AMENDMENT	07/29/2019	CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	* 20	Minus	** 20	= 0		x $50 =	0
	Independent (37 CFR 1.16(h))	* 3	Minus	*** 3	= 0		x $230 =	0
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	0

		(Column 1)		(Column 2)	(Column 3)			
AMENDMENT		CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	*	Minus	**	=		x $0 =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=		x $0 =	
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.	LIE
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".	/LINDA HUMES/
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".	
The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.	

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Docket No.: H0498.70509US02
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	Xiaowei Zhuang
Application No.:	15/329,683
Confirmation No.:	7428
Filed:	January 27, 2017
For:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
Examiner:	F. W. Lu
Art Unit:	1634

Certificate of Electronic Filing under 37 C.F.R. § 1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: July 29, 2019 Electronic Signature for : /Eileen M. MacKenzie/

RESPONSE

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

INTRODUCTORY COMMENTS

In response to the Office Action dated April 4, 2019, reconsideration is respectfully requested in view of the following:

Amendments to the Specification begin on page 2 of this paper.

The Claims are reflected in the listing of claims which begins on page 5 of this paper.

Remarks begin on page 10 of this paper.

AMENDMENTS TO THE SPECIFICATION

Please amend the specification as follows. Additions are indicated by underlining and deletions are indicated by strikeouts and/or double bracketing.

Please amend page 1, lines 4-10 as follows:

This application is a national stage filing of International Patent Application Serial No. PCT/US2015/042556, filed July 29, 2017, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*, which claims the benefit of U.S. Provisional Patent Application Serial No. 62/031,062, filed July 30, 2014, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/142,653, filed April 3, 2015, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; and U.S. Provisional Patent Application Serial No. 62/050,636, filed September 15, 2014, entitled "Probe Library Construction," by Zhuang, *et al.* Each of the above is incorporated herein by reference.

Please amend page 7, lines 5-6 as follows:

Figs. 12A-12C shows schematic descriptions of Hamming distance, in another embodiment of the invention;

Please amend page 7, lines 13-14 as follows:

Figs. 16A-[[16B]]16E show characterization of misidentification rates and calling rates, in one embodiment of the invention;

Please amend page 7, lines 17-20 as follows:

Figs. 18A-18C show a comparison of experiments, in accordance with one embodiment of the invention; ~~and~~

Figs. 19A-19D illustrate decoding and error assessment, in another embodiment of the invention; and

Figs. 20A to 20H illustrate encoding of 130 RNA species and 10 code words as misidentification controls as Codebook 1 and Codebook 2, in another embodiment of the invention.

Please amend the paragraph beginning on line 4 of page 60 as follows:

This example illustrates the measurement of 140 genes with MERFISH using a 16-bit MHD4 Code. To test the feasibility of this error-robust, multiplexed imaging approach, this example uses a 140-gene measurement on human fibroblast cells (IMR90) using a 16-bit MHD4 code to encode 130 RNA species while leaving 10 code words as misidentification controls (Figs. 20A-20H). After each round of hybridization with the fluorescent readout probes, cells were imaged by conventional wide-field imaging with an oblique-incidence illumination geometry. Fluorescent spots corresponding to individual RNAs were clearly detected and were then efficiently extinguished via a brief photobleaching step (Fig. 6A). The sample was stable throughout the 16 rounds of iterative labeling and imaging. The change in the number of fluorescent spots from round to round matched the predicted change based on the relative abundances of RNA species targeted in each round derived from bulk sequencing, and a systematic decreasing trend with increasing number of hybridization rounds was not observed (Fig. 14A). The average brightness of the spots varied from round to round with a standard deviation of 40%, likely due to different binding efficiencies of the readout probes to the different readout sequences on the encoding probes (Fig. 14B). Only a small, systematic decreasing trend in the spot brightness with increasing hybridization rounds was observed, which was on average 4% per round (Fig. 14B).

Please amend the paragraph beginning on line 19 of page 61 as follows:

To test for potential technical bias in these measurements, the same 130 RNAs species were probed with a different MHD4 codebook by shuffling the code words among different RNA species (Figs. 20A to 20H) and changing the encoding probe sequences. Measurements with this alternative

code gave similar misidentification and calling rates (Fig. 17). The copy numbers of individual RNA species per cell measured with these two codebooks showed excellent agreement with a Pearson correlation coefficient of 0.94 (Fig. 6G), indicating that the choice of encoding scheme did not bias the measured counts.

Please amend the paragraph beginning on line 28 of page 64 as follows:

Figs. 20A to 20H show[[s]] two different codebooks for the 140-gene experiments. The specific code words of the 16-bit MHD4 code assigned to each RNA species studied in the two shuffles of the 140-gene experiment. The "Genes" columns contain the name of the gene. The "Codewords" columns contain the specific binary word assigned to each gene.

IN THE CLAIMS

Please replace all prior versions, and listings, of claims in the application with the following list of claims. Additions are indicated by underlining and deletions are indicated by strikeouts and/or double bracketing.

1. (Currently amended) A method, comprising:

associating a plurality of nucleic acid targets in a sample with a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

exposing a sample suspected of containing nucleic acid to a plurality of nucleic acid probes, wherein at least some of the plurality of nucleic acid probes comprises a first portion comprising a target sequence and a second portion comprising one or more read sequences, wherein each read sequence represents a value of a position within the codewords;

for each nucleic acid probe of the plurality of nucleic acid probes, determining binding of the nucleic acid probe[[s]] within the sample;

creating codewords corresponding to the binding of the plurality of nucleic acid probes within the sample, wherein the values of the digits of the codewords are based on the read sequences ~~the binding of~~ present on the nucleic acid probes; ~~and~~

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, either discarding the codeword or applying error correction to the codeword to form a valid codeword, the valid codewords being a plurality of codewords assigned to the plurality of the nucleic acid targets; and

determining a nucleic acid abundance and/or spatial distribution within the sample using the valid codewords corresponding to the binding of the plurality of nucleic acid probes within the sample.

2-6. (Cancelled)

7. (Previously presented) The method of claim 1, wherein the plurality of nucleic acid probes comprises a combinatorial combination of nucleic acid probes with different sequences.

8-9. (Cancelled)

10. (Previously presented) The method of claim 1, wherein at least some of the plurality of nucleic acid probes comprises a first portion comprising a target sequence and a second portion comprising one or more read sequences.

11-17. (Cancelled)

18. (Currently amended) The method of claim 10, wherein the ~~plurality of~~ one or more read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

19-25. (Cancelled)

26. (Currently amended) The method of claim 1, further comprising exposing the sample to a first secondary probe comprising a first signaling entity, the first secondary probe able to bind to some of the read sequences of the nucleic acid probes, and determining binding of the nucleic acid probes by determining the first signaling entity within the sample after removing unbound probe.

27-30. (Cancelled)

31. (Currently amended) The method of claim 26, further comprising exposing the sample to a second secondary probe comprising a second signaling entity, the second secondary probe able to bind to some of read sequences of the nucleic acid probes, and determining binding

of the nucleic acid probes by determining the second signaling entity within the sample after removing unbound probe.

32-33. (Cancelled)

34. (Currently amended) The method of claim 31, further comprising inactivating the first signaling entity or removing the first signaling entity from the sample prior to exposing the sample to the second secondary probe.

35-43. (Cancelled)

44. (Currently amended) The method of claim ~~1~~ 26, comprising determining the centroid of the first signaling entity using an algorithm for determining non-overlapping single emitters.

45-49. (Cancelled)

50. (Previously presented) The method of claim 1, further comprising determining a confidence level for the identified nucleic acid target.

51-70. (Cancelled)

71. (Currently amended) The method of claim 1, comprising determining binding of the nucleic acid probes within the sample at a spatial resolution better than 300 nm.

72-74. (Cancelled)

75. (Previously presented) The method of claim 1, wherein the sample comprises a cell.

76. (Cancelled)

77. (Previously presented) The method of claim 75, wherein the cell is fixed.

78. (Currently amended) The method of any claim 1, comprising determining binding of the nucleic acid probes by imaging at least a portion of the sample using a microscope.

79-94. (Cancelled)

95. (Withdrawn) A composition, comprising:

a plurality of nucleic acid probes, each comprising a first portion comprising a target sequence and a plurality of read sequences, wherein the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

96. (Withdrawn) The composition of claim 95, wherein the plurality of nucleic acid probes defines a code space with a Hamming distance of at least 2.

97-98. (Cancelled)

99. (Withdrawn) The composition of claim 95, wherein the code is a SECDED code.

100-102. (Cancelled)

103. (Withdrawn) The composition of claim 99, wherein only codewords with a constant number of ones are used.

104-121. (Cancelled)

122. (Withdrawn) A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of

codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords; and

forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

123-125. (Cancelled)

126. (Withdrawn) A method, comprising:

exposing nucleic acid probes formed using the method of claim 122 to a sample;

sequentially determining binding of each of a plurality of secondary nucleic probes to the nucleic acid probes within the sample;

determining sample codewords comprising a pattern of bindings of the plurality of secondary nucleic probes to the nucleic acid probes; and

matching the sample codewords to the codewords to determining targets within the sample.

127-128. (Cancelled)

129. (Withdrawn) The method of claim 126, wherein if no match is found between a valid codeword and the sample codeword, applying error correction to the sample codeword to identified a matched codeword.

130-176. (Cancelled)

REMARKS

Claim 1 has been amended as shown above. Claim 18 has been amended to recite "the one or more read sequences." Claims 26 and 31 have both been amended to recite "after removing unbound probe." Claim 34 has been amended to recite "or removing the first signaling entity from the sample." Claim 71 has been amended to recite "spatial resolution." Claim 78 has been amended to recite "using a microscope."

In addition, minor amendments have been made to the specification to identify the figures and update the priority claim, as noted above.

No new matter has been added. Claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 are currently pending for examination. Claims 95, 96, 99 103, 122, 126, and 129 have been withdrawn.

Specification

The specification has been objected to because of various informalities in the figure numbering.

These have been corrected as noted above. Accordingly, withdrawal of this objection is respectfully requested.

Claim Objections

Claim 18 has been objected to because "the plurality of read sequences" is not consistent with claim 10.

Claim 18 has been amended to recite "the one or more read sequences." Accordingly, withdrawal of this objection is respectfully requested.

Rejections under 35 U.S.C. § 112, ¶ 1

Claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 have been rejected under 35 U.S.C. § 112, ¶ 1, because the specification, while being enabling for exposing a sample to a plurality of nucleic acid probes, does not reasonably provide enablement for performing the methods recited in claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78.

The accuracy of this rejection under *In re Wands* is not conceded, including the level of one of ordinary skill in the art. Furthermore, a discussion of each of the *Wands* factors with respect to each of the specific claim rejections was not provided, which is improper, as the Patent Office must set forth its reasoning regarding each of the *Wands* factors with particularity.

However, solely to further the patent application process, claim 1 has been amended to recite, *inter alia*, a sample suspected of containing nucleic acid, codewords that comprise a number of positions and values for each position and form an error-checking and/or error-correcting code space, and discarding the codeword or applying error correction to the codeword to form a valid codeword if no match is found between the codeword to a valid codeword. In addition, claims 26 and 31 have both been amended to recite "after removing unbound probe"; claim 34 has been amended to recite "or removing the first signaling entity from the sample"; and claim 78 has been amended to recite "using a microscope."

It is believed that these amendments will address the various enablement rejections set forth in the Office Action, including the rejections of independent claim 1 and dependent claims 26, 31, 34, 44, and 78. Accordingly, withdrawal of these rejections is respectfully requested.

Rejections under 35 U.S.C. § 112, ¶ 2

Claim 71 has been rejected under 35 U.S.C. § 112, ¶ 2, for being indefinite for failing to particularly point out and distinctly claim the subject matter which the applicant regards as the invention.

Claim 71 has been amended to clarify that the resolution of better than 300 nm is spatial resolution. However, it is noted that determining spatial resolution of an entity is not limited to any one particular device or technique. Any device able to determine such spatial resolutions may be used in claim 71. As examples, a variety of techniques are provided in the specification, many of which are commercially available. See, e.g., page 28, line 3 to page 29, line 12. One of ordinary skill in the art would accordingly have had no undue difficulty in determining binding within a sample at a spatial resolution better than 300 nm. Thus, withdrawal of this rejection is respectfully requested.

CONCLUSION

In view of the foregoing remarks, this application should now be in condition for allowance. A notice to this effect is respectfully requested. If the Examiner believes, after the foregoing remarks, that the application is not in condition for allowance, the Examiner is requested to call the undersigned at 617-646-8000.

If this response is not considered timely filed and if a suitable request for an extension of time is otherwise absent, any necessary extension of time is hereby requested. If there is a fee occasioned by this response, including an extension fee, the Director is hereby authorized to charge any deficiency or credit any overpayment in the fees filed, asserted to be filed or which should have been filed herewith to Deposit Account No. 23/2825, under Docket No. H0498.70509US02, from which the undersigned is authorized to draw.

Dated: July 29, 2019

Respectfully submitted,

Electronic signature: /Tani Chen/
Tani Chen, Sc.D.
Registration No.: 52,728
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

RELATED APPLICATIONS

This application is a national stage filing of International Patent Application Serial No. PCT/US2015/042556, filed July 29, 2017, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*, which claims the benefit of U.S. Provisional Patent Application Serial No. 62/031,062, filed July 30, 2014, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/142,653, filed April 3, 2015, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; and U.S. Provisional Patent Application Serial No. 62/050,636, filed September 15, 2014, entitled "Probe Library Construction," by Zhuang, *et al.* Each of the above is incorporated herein by reference.

GOVERNMENT FUNDING

This invention was made with government support under Grant No. GM096450 awarded by National Institutes of Health. The government has certain rights in the invention.

FIELD

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined.

BACKGROUND

Single-molecule fluorescent in situ hybridization (smFISH) is a powerful method for detecting individual mRNA molecules in cells. The high detection efficiency and large dynamic range of this method provides exquisite detail into the expression state, spatial distribution within cells and intact tissues, and variation among cells of individual mRNAs. Such approaches have been essential to many recent insights into understanding gene regulation and expression. A fundamental limitation of smFISH, however, is its low throughput, typically only a few genes at a time. This low throughput is due to a lack of distinguishable probes with which to label cells and the cost of producing large amounts of labeled probe required for high efficiency staining. Thus, improvements in detecting mRNA molecules are needed.

SUMMARY

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. The subject matter of the present invention

involves, in some cases, interrelated products, alternative solutions to a particular problem, and/or a plurality of different uses of one or more systems and/or articles.

In one aspect, the present invention is generally directed to a composition. According to one set of embodiments, the composition comprises a plurality of nucleic acid probes, at least some of which comprise a first portion comprising a target sequence and a plurality of read sequences. In some cases, each comprises a first portion comprising a target sequence and a plurality of read sequences. In some embodiments, the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

In another aspect, the present invention is generally directed to a method. In one set of embodiments, the method includes acts of exposing a sample to a plurality of nucleic acid probes; for each of the nucleic acid probes, determining binding of the nucleic acid probes within the sample; creating codewords based on the binding of the nucleic acid probes; and for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

The method, in another set of embodiments, includes acts of exposing a sample to a plurality of nucleic acid probes, wherein the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences, and wherein at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences; and for each of the nucleic acid probes, determining binding of the target sequences of the nucleic acid probes within the sample.

In yet another set of embodiments, the method includes acts of exposing a sample to a plurality of primary nucleic acid probes (also called encoding probes); exposing the plurality of primary nucleic acid probes to a sequence of secondary nucleic acid probes (also called readout probes) and determining fluorescence of each of the secondary nucleic acid probes within the sample; creating codewords based on fluorescence of the secondary nucleic acid probes; and for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

In one set of embodiments, the method includes acts of exposing a plurality of primary nucleic acid probes to a sample; and exposing the plurality of nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary probes within the sample. In some embodiments, at least some of the plurality of secondary nucleic acid probes comprises distinguishable secondary nucleic acid probes

formed from combinatorial combination of one or more read sequences (or readout probe sequences) taken from a plurality of read sequences (or readout probe sequences).

In another set of embodiments, the method comprises acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a precision better than 500 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a precision better than 500 nm.

In another set of embodiments, the method comprises acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a resolution better than 100 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a resolution better than 100 nm.

The method, in yet another set of embodiments, includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity using a super-resolution imaging technique, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In certain embodiments, the method comprises acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords; and forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

In addition, in one set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes each comprising a target sequence; and forming groups comprising the

plurality of nucleic acid probes such that each group of nucleic acid probes corresponds to at least one common value of a position within the codewords.

In another set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions that is less than the number of targets, and wherein each codeword is associated with a single target, associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords, and forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

The method, in still another set of embodiments, includes acts of exposing a plurality of nucleic acid probes to a cell, exposing the plurality of nucleic acid probes to a sequence of secondary probes and determining fluorescence of each of the secondary probes within the cell, and based on the sequence of fluorescence of each of the secondary probes, determining nucleic acids within the cell.

In another set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes each comprising a target sequence; and forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes correspond to at least one common value of a position within the codewords.

In yet another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity using a super-resolution imaging technique, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In still another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a precision better than 500 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a precision better than 500 nm.

In still another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a resolution better than 100 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In still another set of embodiments, the method includes acts of associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code; associating unique read sequences with each possible value of each position in the codewords, wherein the read sequences are taken from a set of orthogonal sequences, which have limited homology with one another and with the nucleic acid species in a sample; forming a plurality of primary nucleic acid probes each comprising a target sequence that uniquely binds to a nucleic acid target and one or more read sequences; forming a plurality of secondary nucleic acid probes comprising a signaling entity and a sequence that is complementary to one of the read sequences; exposing a sample to the primary nucleic acid probes such that the nucleic acid probes hybridize to the nucleic acid targets in the sample; exposing the primary nucleic acid probes in the sample to a secondary nucleic acid probe such that the secondary nucleic acid probe hybridizes to the read sequence on at least some of the primary nucleic acid probes; imaging the sample; and repeating the exposing and imaging steps one or more times, using a different secondary nucleic acid probe for at least some of the repetitions.

The method, according to yet another set of embodiments, includes acts of associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes comprising a signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets; grouping the nucleic acid probes into a plurality of probe pools, wherein each of the probe pools corresponds to a specific value of a unique position within the codewords; exposing a sample to one of the probe pools; imaging the sample; and repeating the exposing and imaging steps one or more times, using a different probe pool for at least some of the repetitions.

In another aspect, the present invention encompasses methods of making one or more of the embodiments described herein. In still another aspect, the present invention encompasses methods of using one or more of the embodiments described herein.

Other advantages and novel features of the present invention will become apparent from the following detailed description of various non-limiting embodiments of the invention when considered in conjunction with the accompanying figures. In cases where the present specification and a document incorporated by reference include conflicting and/or inconsistent disclosure, the present specification shall control. If two or more documents incorporated by reference include conflicting and/or inconsistent disclosure with respect to each other, then the document having the later effective date shall control.

BRIEF DESCRIPTION OF THE DRAWINGS

Non-limiting embodiments of the present invention will be described by way of example with reference to the accompanying figures, which are schematic and are not intended to be drawn to scale. In the figures, each identical or nearly identical component illustrated is typically represented by a single numeral. For purposes of clarity, not every component is labeled in every figure, nor is every component of each embodiment of the invention shown where illustration is not necessary to allow those of ordinary skill in the art to understand the invention. In the figures:

Figs. 1A-1C illustrates an encoding scheme for nucleic acid probes, in certain embodiments of the invention;

Figs. 2A-2G illustrates determining of mRNAs in a cell, in some embodiments of the invention;

Figs. 3A-3B illustrate the determination of nucleic acids, in accordance with various embodiments of the invention;

Figs. 4A-4B is a non-limiting example of multiple read sequences distributed in a population of different nucleic acid probes, in accordance with certain embodiments of the invention;

Figs. 5A-5E illustrate the determination of nucleic acids, in accordance with another embodiment of the invention;

Figs. 6A-6H illustrate simultaneous determination of multiple nucleic acid species in cells, in certain embodiments of the invention;

Figs. 7A-7E show expression noise of genes and co-variation of expression between different genes determined in accordance with some embodiments of the invention;

Figs. 8A-8E illustrate spatial distribution of RNAs in cells determined in accordance with one embodiment of the invention;

Figs. 9A-9C illustrate simultaneous determination of multiple nucleic acid species in cells, in another embodiment of the invention;

Figs. 10A-10B show expression between different genes determined in accordance with yet another embodiment of the invention;

Fig. 11 is a schematic description of combinatorial labeling, in accordance with another embodiment of the invention;

Figs. 12A-12C shows schematic descriptions of Hamming distance, in another embodiment of the invention;

Fig. 13 illustrates the production of a library of probes, in still another embodiment of the invention;

Figs. 14A-14B illustrate fluorescent spot determinations, in another embodiment of the invention;

Figs. 15A-15B illustrate error correction facilitates RNA detection, in yet another embodiment of the invention;

Figs. 16A-[[16B]]16E show characterization of misidentification rates and calling rates, in one embodiment of the invention;

Figs. 17A-17D show characterization of misidentification rates and calling rates, in another embodiment of the invention;

Figs. 18A-18C show a comparison of experiments, in accordance with one embodiment of the invention; ~~and~~

Figs. 19A-19D illustrate decoding and error assessment, in another embodiment of the invention; and

Figs. 20A to 20H illustrate encoding of 130 RNA species and 10 code words as misidentification controls as Codebook 1 and Codebook 2, in another embodiment of the invention.

BRIEF DESCRIPTION OF THE SEQUENCES

SEQ ID NO: 1 is:

GTTGGCGACGAAAGCACTGCGATTGGAACCGTCCCAAGCGTTGCGCTTAATGGA TCATCAATTTTGTCTCACTACGACGGTCAATCGCGCTGCATACTTGCGTCGGTCG GACAAACGAGG;

SEQ ID NO: 2 is CGCAACGCTTGGGACGGTTCCAATCGGATC;

SEQ ID NO: 3 is CGAATGCTCTGGCCTCGAACGAACGATAGC;

SEQ ID NO: 4 is ACAAATCCGACCAGATCGGACGATCATGGG;

SEQ ID NO: 5 is CAAGTATGCAGCGCGATTGACCGTCTCGTT;

SEQ ID NO: 6 is TGCGTCGTCTGGCTAGCACGGCACGCAAAT;

SEQ ID NO: 7 is AAGTCGTACGCCGATGCGCAGCAATTCACT;

SEQ ID NO: 8 is CGAAACATCGGCCACGGTCCCGTTGAACTT;

SEQ ID NO: 9 is ACGAATCCACCGTCCAGCGCGTCAAACAGA;

SEQ ID NO: 10 is CGCGAAATCCCCGTAACGAGCGTCCCTTGC;

SEQ ID NO: 11 is GCATGAGTTGCCTGGCGTTGCGACGACTAA;

SEQ ID NO: 12 is CCGTCGTCTCCGGTCCACCGTTGCGCTTAC;

SEQ ID NO: 13 is GGCCAATGGCCCAGGTCCGTCACGCAATTT;

SEQ ID NO: 14 is TTGATCGAATCGGAGCGTAGCGGAATCTGC;

SEQ ID NO: 15 is CGCGCGGATCCGCTTGTCGGGAACGGATAC;

SEQ ID NO: 16 is GCCTCGATTACGACGGATGTAATTCGGCCG;

SEQ ID NO: 17 is GCCCGTATTCCCGCTTGCGAGTAGGGCAAT

SEQ ID NO: 18 is GTTGGTCGGCACTTGGGTGC;

SEQ ID NO: 19 is CGATGCGCCAATTCCGGTTC;

SEQ ID NO: 20 is CGCGGGCTATATGCGAACCG;

SEQ ID NO: 21 is
TAATACGACTCACTATAGGGAAAGCCGGTTCATCCGGTGG;

SEQ ID NO: 22 is
TAATACGACTCACTATAGGGTGATCATCGCTCGCGGGTTG;

SEQ ID NO: 23 is
TAATACGACTCACTATAGGGCGTGGAGGGCATACAACGC;

SEQ ID NO: 24 is CGCAACGCTTGGGACGGTTCCAATCGGATC/3Cy5Sp/;

SEQ ID NO: 25 is CGAATGCTCTGGCCTCGAACGAACGATAGC/3Cy5Sp/;

SEQ ID NO: 26 is ACAAATCCGACCAGATCGGACGATCATGGG/3Cy5Sp/;

SEQ ID NO: 27 is CAAGTATGCAGCGCGATTGACCGTCTCGTT/3Cy5Sp/;

SEQ ID NO: 28 is GCGGGAAGCACGTGGATTAGGGCATCGACC/3Cy5Sp/;

SEQ ID NO: 29 is AAGTCGTACGCCGATGCGCAGCAATTCACT/3Cy5Sp/;

SEQ ID NO: 30 is CGAAACATCGGCCACGGTCCCGTTGAACTT/3Cy5Sp/;

SEQ ID NO: 31 is ACGAATCCACCGTCCAGCGCGTCAAACAGA/3Cy5Sp/;

SEQ ID NO: 32 is CGCGAAATCCCCGTAACGAGCGTCCCTTGC/3Cy5Sp/;

SEQ ID NO: 33 is GCATGAGTTGCCTGGCGTTGCGACGACTAA/3Cy5Sp/;

SEQ ID NO: 34 is CCGTCGTCTCCGGTCCACCGTTGCGCTTAC/3Cy5Sp/;

SEQ ID NO: 35 is GGCCAATGGCCCAGGTCCGTCACGCAATTT/3Cy5Sp/;
SEQ ID NO: 36 is TTGATCGAATCGGAGCGTAGCGGAATCTGC/3Cy5Sp/;
SEQ ID NO: 37 is CGCGCGGATCCGCTTGTCGGGAACGGATAC/3Cy5Sp/;
SEQ ID NO: 38 is GCCTCGATTACGACGGATGTAATTCGGCCG/3Cy5Sp/; and
SEQ ID NO: 39 is GCCCGTATTCCCGCTTGCGAGTAGGGCAAT/3Cy5Sp/

.

DETAILED DESCRIPTION

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. Certain embodiments are directed to determining nucleic acids, such as mRNA, within cells at relatively high resolutions. In some embodiments, a plurality of nucleic acid probes may be applied to a sample, and their binding within the sample determined, e.g., using fluorescence, to determine locations of the nucleic acid probes within the sample. In some embodiments, codewords may be based on the binding of the plurality of nucleic acid probes, and in some cases, the codewords may define an error-correcting code to reduce or prevent misidentification of the nucleic acids. In certain cases, a relatively large number of different targets may be identified using a relatively small number of labels, e.g., by using various combinatorial approaches.

Two example approaches are now discussed. It should be understood, however, that these are presented by way of explanation and not limitation; other aspects and embodiments are discussed in further detail herein. In one example method, primary probes (also called encoding probes) and secondary probes (also called readout probes) are used, where the primary probes encode "codewords" and bind to target nucleic acids in the sample, and the secondary probes are used to read out the codewords from the primary probes. In another example method, a plurality of different primary probes containing codewords are divided into as many separate pools as there are positions in the codewords, such that each primary probe pool corresponds to a certain value in a certain position of the codewords (e.g., a "one" in the first position as in "1001").

The first example is now described with respect to Fig. 3A. As will be discussed in more detail below, in other embodiments, other configurations may be used as well. In this first example, a series of nucleic acid probes are used to determine nucleic acids within a cell or other sample, e.g., qualitatively or quantitatively. For example, nucleic acids may be identified as being present or absent, and/or the numbers or concentrations of certain nucleic acids may be determined within the cell or other sample. In some cases, the positions of the

probes within the cell or other sample may be determined at relatively high resolutions, and in some cases, at resolutions better than the wavelength of visible light.

This example is generally directed to spatially detecting nucleic acids within a cell or other sample, e.g., at relatively high resolutions. For example, the nucleic acids may be mRNAs, or other nucleic acids described herein. In one set of embodiments, the nucleic acids within the cell may be determined by delivering or applying nucleic acid probes to the cell. In some cases, by using combinatorial approaches, a relatively large number of nucleic acids may be determined using a relatively small number of different labels on the nucleic acid probes. Thus, for example, a relatively small number of experiments may be used to determine a relatively large number of nucleic acids in a sample, e.g., due to simultaneous binding of the nucleic acid probes to different nucleic acids in the sample.

In one set of embodiments, a population of primary nucleic acid probes are applied to the cell (or other sample) that is able to bind nucleic acids suspected of being present within the cell. Afterwards, sequentially, secondary nucleic acid probes that can bind to or otherwise interact with some of the primary nucleic acids are added and determined, e.g., using imaging techniques such as fluorescence microscopy (e.g., conventional fluorescence microscopy), STORM (stochastic optical reconstruction microscopy) or other imaging techniques. After imaging, the secondary nucleic acid probes are inactivated or removed, and a different secondary nucleic acid probe is added to the sample. This may be repeated multiple times with multiple different secondary nucleic acid probes. The pattern of binding of the various secondary nucleic acid probes may be used to determine the primary nucleic acid probes at locations within the cell or other sample, which can be used to determine mRNA or other nucleic acids that are present.

For instance, as is shown in Fig. 3A, a population of nucleic acids 10 within a cell (represented here by nucleic acids 11, 12, and 13) may be exposed to a population of primary nucleic acid probes 20, including probes 21 and 22. The primary nucleic acid probes may contain, for instance, a target sequence that can recognize a nucleic acid (e.g., a sequence within nucleic acid 11). Probes 21 and 22 may contain the same or different targeting sequences, which may bind to or hybridize with the same or different nucleic acids. As an example, as is shown in Fig. 3A, probe 21 contains a first targeting sequence 25 which targets the probe to nucleic acid 11, while probe 22 contains a second targeting sequence 26, not identical to the first targeting sequence 25 and which targets the probe to nucleic acid 12. The target sequence may be substantially complementary to at least a portion of a target

nucleic acid, and enough of the target sequence may be present such that specific binding of the nucleic acid probe to the target nucleic acid can occur.

Primary nucleic acid probes 20 may also contain one or more "read" sequences. Two such read sequences are used in this example, although in other embodiments, there may be one, three, four, or more read sequences present within a primary nucleic acid probe. The read sequences may all independently be the same or different. In addition, in one set of embodiments, different nucleic acid probes may use one or more common read sequences. For example, more than one read sequence may be combinatorially combined on different nucleic acid probes, thereby producing a relatively large number of different nucleic acid probes that can be separately identified, even though only a relatively small number of read sequences are used. Thus, for example, in Fig. 3A, probe 21 contains read sequences 27 and 29, while probe 22 contains read sequences 27 and 28, where the two read sequences 27 are identical, and different from read sequences 28 and 29.

After primary nucleic acid probes 20 have been introduced to the sample and allowed to interact with nucleic acids 11, 12, and 13, one or more secondary nucleic acid probes 30 may be applied to the sample to determine the primary nucleic acid probes. The secondary nucleic acid probes may contain a recognition sequence able to recognize one of the read sequences present within the population of primary nucleic acid probes. For instance, the recognition sequence may be substantially complementary to at least a portion of the read sequence, such that the secondary nucleic acid probe is able to bind to or hybridize with corresponding primary nucleic acid probe. For instance, in this example, recognition sequence 35 is able to recognize read sequence 27. In addition, the secondary nucleic acid probes may contain one or more signaling entities 33. For example, a signaling entity may be a fluorescent entity attached to the probe, or a certain sequence of nucleic acids that can be determined in some fashion. More than one secondary sequence may be used, e.g., sequentially. For example, as shown in this figure, the initial secondary probe 30 may be removed (e.g., as discussed below) and a new secondary probe 31 may be added, containing recognition sequence 36 able to recognize read sequence 28 and one or more signaling entities 33. This may also be repeated multiple times, e.g., to determine read sequence 29 or other read sequences that may be present.

The location of the secondary nucleic acid probes 30, 31, etc. may be determined by determining signaling entity 33. For example, if the signaling entity is fluorescent, then fluorescence microscopy can be used to determine the signaling entity. In some embodiments, imaging of a sample to determine the signaling entity may be used at relatively

high resolutions, and in some cases, super-resolution imaging techniques (e.g., resolutions better than the wavelength of visible light or the diffraction limit of light) may be used. Examples of super-resolution imaging techniques include STORM, or other techniques as discussed herein. In some cases, e.g., with certain super-resolution imaging techniques such as STORM, more than one image of the sample may be acquired.

More than one type of secondary nucleic acid probe may be applied to a cell or other sample. For example, a first secondary nucleic acid probe may be applied that can recognize a first read sequence, then it or its attached signaling entity may be inactivated or removed, and a second secondary nucleic acid probe may be applied that can recognize a second read sequence. This process may be repeated multiple times, each with a different secondary nucleic acid probe, e.g., to determine the read sequences that were present in the various primary nucleic acid probes. Thus, primary nucleic acids within the sample can be determined on the basis of the binding pattern of secondary nucleic acid probes.

For example, a first location within the cell or other sample may exhibit binding of a first secondary probe and a third secondary probe, but not the binding of a second or a fourth secondary probe, while a second location may exhibit a different pattern of binding of various secondary probes. The primary nucleic acid probe that the secondary probes are able to bind to or hybridize with may be determined by considering the pattern of binding of various secondary probes. For instance, referring to Fig. 3A, if a first secondary probe is able to determine read sequence 27, a second secondary probe is able to determine read sequence 28, and a third secondary probe is able to determine read sequence 29, then primary nucleic acid 25 may be determined through the binding of the first and third secondary probes (but not the second secondary probe), while primary nucleic acid 26 may be determined through the binding of the first and second secondary probes (but not the third secondary probe). Similarly, if it is known that first probe 21 contains target sequence 25 while second probe 22 contains target sequence 26, then nucleic acids 11 and 12 may also be determined within the sample, e.g., spatially, based on the binding pattern of the various secondary nucleic acid probes. In addition, it should be noted that due to the presence of more than one read sequence on the primary nucleic acid probes, even though first probe 21 and second probe 22 contains a common read sequence (read sequence 27), these probes may be distinguished in the sample due to the different binding patterns of the various secondary nucleic acid probes.

In certain embodiments, this pattern of binding or hybridization of the secondary nucleic acid probes may be converted into a "codeword." In this example, for instance, the codewords are "101" and "110" for first probe 21 and second probe 22, respectively, where a

value of 1 represents binding and a value of 0 represents no binding. The codewords may also have longer lengths in other embodiments; only three probes are shown here for clarity purposes only. A codeword can be directly related to a specific target nucleic acid sequence of the primary nucleic acid probe. Accordingly, different primary nucleic acid probes may match certain codewords, which can then be used to identify the different targets of the primary nucleic acid probes based on the binding patterns of the secondary probes, even if in some cases, there is overlap in the read sequences of different secondary probes, e.g., as was shown in Fig. 3A. However, if no binding is evident (e.g., for nucleic acid 13), then the codeword would be "000" in this example.

The values in each codeword can also be assigned in different fashions in some embodiments. For example, a value of 0 could represent binding while a value of 1 represents no binding. Similarly, a value of 1 could represent binding of a secondary nucleic acid probe with one type of signaling entity while a value of 0 could represent binding of a secondary nucleic acid probe with another type of distinguishable signaling entity. These signaling entities could be distinguished, for example, via different colors of fluorescence. In some cases, values in codewords need not be confined to 0 and 1. The values could also be drawn from larger alphabets, such as ternary (e.g., 0, 1, and 2) or quaternary (e.g., 0, 1, 2, and 3) systems. Each different value could, for example, be represented by a different distinguishable signaling entity, including (in some cases) one value that may be represented by the absence of signal.

The codewords for each target may be assigned sequentially, or may be assigned at random. For instance, referring to Fig. 3A, a first nucleic acid target may be assigned to 101, while a second nucleic acid target may be assigned to 110. In addition, in some embodiments, the codewords may be assigned using an error-detection system or an error-correcting system, such as a Hamming system, a Golay code, or an extended Hamming system (or a SECDED system, i.e., single error correction, double error detection). Generally speaking, such systems can be used to identify where errors have occurred, and in some cases, such systems can also be used to correct the errors and determine what the correct codeword should have been. For example, a codeword such as 001 may be detected as invalid and corrected using such a system to 101, e.g., if 001 is not previously assigned to a different target sequence. A variety of different error-correcting codes can be used, many of which have previously been developed for use within the computer industry; however, such error-correcting systems have not typically been used within biological systems. Additional examples of such error-correcting codes are discussed in more detail below.

It should also be understood that all possible codewords in a code need not be used in some cases. For example, in some embodiments, codewords that are not used can serve as negative controls. Similarly, in some embodiments, some codewords can be left out because they are more prone to errors in measurement than other codewords. For example, in some implementations, reading a codeword with more values of '1' might be more error-prone that reading a codeword with fewer values of '1.'

It should be understood that the above description is an example of one embodiment of the invention, and that primary and secondary nucleic acid probes are not necessary in all embodiments. For example, in some embodiments, a series of nucleic acid probes containing signaling entities are used to determine nucleic acids within a cell or other sample, without necessarily requiring secondary probes.

For example, turning now to Fig. 3B, nucleic acids 11, 12, and 13 are exposed to different rounds of probes 21, 22, 23, 24, etc. in this example. These probes may each contain a target sequence that can recognize a nucleic acid (e.g., a sequence within nucleic acid 11 or 12). These probes may each target the same nucleic acid, but different regions of the nucleic acid. In addition, some or all of the probes may contain one or more signaling entities, e.g., signaling entity 29 on probe 21. For example, the signaling entity may be a fluorescent entity attached to the probe, or a certain sequence of nucleic acids that can be determined in some fashion.

The first round of probes (e.g. probe 21 and probe 22) may be applied to the cell or other sample. Probe 21 may be allowed to bind to nucleic acid 11 via target sequence 25. Such binding can be determined by determining signaling entity 29. For example, if the signaling entity is fluorescent, then fluorescence microscopy can be used to determine the signaling entity, e.g., spatially within the cell or other sample. In some but not all embodiments, imaging of a sample to determine the signaling entity may be used at relatively high resolutions, and in some cases, super-resolution imaging techniques may be used. Other, different probes may be present as well; for instance, probe 22 containing target sequence 26 may bind to nucleic acid 12, and be determined via signaling entity 29 within probe 22. These may occur, e.g., sequentially or simultaneously. Optionally, probes 21 and 22 may also be removed or inactivated, e.g., between application of different rounds of probes.

Next, a second round of probes (e.g., probe 23) is applied to the sample. In this example, probe 23 is able to bind to nucleic acid 11 via a targeting region, although there is no probe in the second round that is able to bind to nucleic acid 12. Binding of the probes is

allowed to occur as discussed above, and determination of binding may occur via signaling entities. These signaling entities may be the same or different as from the first round of probes. This process may be repeated any number of times with different probes. For example, as is shown in Fig. 3B, round 2 contains probes able to bind to nucleic acid 11, while round 3 contains probes able to bind to nucleic acid 12.

In certain embodiments, each round of binding or hybridization of nucleic acid probes may be converted into a "codeword." In this example, using probes 21, 22, 23, and 24, the codewords 101 or 110 could be formed, where 1 represents binding and 0 represents no binding and the first position corresponds to the binding of probes 21 or 22 while the second position corresponds to the binding of probes 22, and the third position corresponds to the binding of probe 24. A codeword of 000 would represent no binding, e.g., as shown with nucleic acid 13 in this example. A codeword can be directly related to a specific target nucleic acid sequence of the nucleic acid probes, by designing appropriate nucleic acid probes. Thus, for example, 110 may correspond to a first target nucleic acid 12 (e.g., the first and second round of nucleic acid probes containing probes able to target nucleic acid 11, and these probes may target the same or different regions of nucleic acid 11) while 101 may correspond to a second target nucleic acid (e.g., the first and third round of nucleic acid probes containing probes able to target nucleic acid 12, and these probes may target the same or different regions of nucleic acid 12). In addition, it should be noted that each round of probes may contain the same, or different signaling entities as other probes in the same round, and/or other probes in different rounds. For instance, in one set of embodiments, only one signaling entity is used in all of the rounds of probes.

Similar to the above, the codewords for each target may be assigned sequentially, or may be assigned at random. The codewords may be assigned within a code space in some embodiments using an error-detection or an error-correcting system, such as a Hamming system, a Golay code, or an extended Hamming system or a SECDED system (single error correction, double error detection). Generally speaking, such error-correction systems can be used to identify where errors have occurred, and in some cases, such systems can also be used to correct the errors and determine what the correct codeword should have been.

Similar to the above, the values at each position in the codeword can be arbitrarily assigned in certain embodiments to binding or non-binding of probes that contain more than one distinguishable signaling entity.

In some cases, the nucleic acid probes may be formed into "pools" or groups of nucleic acids that share a common feature. For example, probes to all targets with codewords

that contain a 1 in the first position, e.g. 110 and 101 but not 011, may comprise one pool while probes to all targets that contain a 1 in the second position, e.g. 110 and 011 but not 101, may comprise another pool. See also Fig. 1C. In some cases, a nucleic acid probe may be a member of more than one group or pool. Members of a nucleic acid pool may also contain features in addition to target sequences, read sequences, and or signaling entities that allow them to be distinguished from other groups. These features may be short nucleic acid sequences that are used for the amplification, production, or separation of these sequences. The nucleic acid probes of each group may be applied to a sample, e.g., sequentially, as discussed herein.

Thus, in some aspects, the present invention is generally directed to systems and methods for determining nucleic acids within a cell or other sample. The sample may include a cell culture, a suspension of cells, a biological tissue, a biopsy, an organism, or the like. The sample may also be cell-free but nevertheless contain nucleic acids. If the sample contains a cell, the cell may be a human cell, or any other suitable cell, e.g., a mammalian cell, a fish cell, an insect cell, a plant cell, or the like. More than one cell may be present in some cases.

The nucleic acids to be determined may be, for example, DNA, RNA, or other nucleic acids that are present within a cell (or other sample). The nucleic acids may be endogenous to the cell, or added to the cell. For instance, the nucleic acid may be viral, or artificially created. In some cases, the nucleic acid to be determined may be expressed by the cell. The nucleic acid is RNA in some embodiments. The RNA may be coding and/or non-coding RNA. Non-limiting examples of RNA that may be studied within the cell include mRNA, siRNA, rRNA, miRNA, tRNA, lncRNA, snoRNAs, snRNAs, exRNAs, piRNAs, or the like.

In some cases, a significant portion of the nucleic acid within the cell may be studied. For instance, in some cases, enough of the RNA present within a cell may be determined so as to produce a partial or complete transcriptome of the cell. In some cases, at least 4 types of mRNAs are determined within a cell, and in some cases, at least 3, at least 4, at least 7, at least 8, at least 12, at least 14, at least 15, at least 16, at least 22, at least 30, at least 31, at least 32, at least 50, at least 63, at least 64, at least 72, at least 75, at least 100, at least 127, at least 128, at least 140, at least 255, at least 256, at least 500, at least 1,000, at least 1,500, at least 2,000, at least 2,500, at least 3,000, at least 4,000, at least 5,000, at least 7,500, at least 10,000, at least 12,000, at least 15,000, at least 20,000, at least 25,000, at least 30,000, at least 40,000, at least 50,000, at least 75,000, or at least 100,000 types of mRNAs may be determined within a cell.

In some cases, the transcriptome of a cell may be determined. It should be understood that the transriptome generally encompasses all RNA molecules produced within a cell, not just mRNA. Thus, for instance, the transcriptome may also include rRNA, tRNA, siRNA, etc. In some embodiments, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or 100% of the transcriptome of a cell may be determined.

The determination of one or more nucleic acids within the cell or other sample may be qualitative and/or quantitative. In addition, the determination may also be spatial, e.g., the position of the nucleic acid within the cell or other sample may be determined in two or three dimensions. In some embodiments, the positions, number, and/or concentrations of nucleic acids within the cell (or other sample) may be determined.

In some cases, a significant portion of the genome of a cell may be determined. The determined genomic segments may be continuous or interspersed on the genome. For example, in some cases, at least 4 genomic segments are determined within a cell, and in some cases, at least 3, at least 4, at least 7, at least 8, at least 12, at least 14, at least 15, at least 16, at least 22, at least 30, at least 31, at least 32, at least 50, at least 63, at least 64, at least 72, at least 75, at least 100, at least 127, at least 128, at least 140, at least 255, at least 256, at least 500, at least 1,000, at least 1,500, at least 2,000, at least 2,500, at least 3,000, at least 4,000, at least 5,000, at least 7,500, at least 10,000, at least 12,000, at least 15,000, at least 20,000, at least 25,000, at least 30,000, at least 40,000, at least 50,000, at least 75,000, or at least 100,000 genomic segments may be determined within a cell.

In some cases, the entire genome of a cell may be determined. It should be understood that the genome generally encompasses all DNA molecules produced within a cell, not just chromosome DNA. Thus, for instance, the genome may also include, in some cases, mitochondria DNA, chloroplast DNA, plasmid DNA, etc. In some embodiments, at least about 5%, at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or 100% of the genome of a cell may be determined.

As discussed herein, a variety of nucleic acid probes may be used to determine one or more nucleic acids within a cell or other sample. The probes may comprise nucleic acids (or entities that can hybridize to a nucleic acid, e.g., specifically) such as DNA, RNA, LNA (locked nucleic acids), PNA (peptide nucleic acids), or combinations thereof. In some cases,

additional components may also be present within the nucleic acid probes, e.g., as discussed below. Any suitable method may be used to introduce nucleic acid probes into a cell.

For example, in some embodiments, the cell is fixed prior to introducing the nucleic acid probes, e.g., to preserve the positions of the nucleic acids within the cell. Techniques for fixing cells are known to those of ordinary skill in the art. As non-limiting examples, a cell may be fixed using chemicals such as formaldehyde, paraformaldehyde, glutaraldehyde, ethanol, methanol, acetone, acetic acid, or the like. In one embodiment, a cell may be fixed using Hepes-glutamic acid buffer-mediated organic solvent (HOPE).

The nucleic acid probes may be introduced into the cell (or other sample) using any suitable method. In some cases, the cell may be sufficiently permeabilized such that the nucleic acid probes may be introduced into the cell by flowing a fluid containing the nucleic acid probes around the cells. In some cases, the cells may be sufficiently permeabilized as part of a fixation process; in other embodiments, cells may be permeabilized by exposure to certain chemicals such as ethanol, methanol, Triton, or the like. In addition, in some embodiments, techniques such as electroporation or microinjection may be used to introduce nucleic acid probes into a cell or other sample.

Certain aspects of the present invention are generally directed to nucleic acid probes that are introduced into a cell (or other sample). The probes may comprise any of a variety of entities that can hybridize to a nucleic acid, typically by Watson-Crick base pairing, such as DNA, RNA, LNA, PNA, etc., depending on the application. The nucleic acid probe typically contains a target sequence that is able to bind to at least a portion of a target nucleic acid, in some cases specifically. When introduced into a cell or other system, the target system may be able to bind to a specific target nucleic acid (e.g., an mRNA, or other nucleic acids as discussed herein). In some cases, the nucleic acid probes may be determined using signaling entities (e.g., as discussed below), and/or by using secondary nucleic acid probes able to bind to the nucleic acid probes (i.e., to primary nucleic acid probes). The determination of such nucleic acid probes is discussed in detail below.

In some cases, more than one type of (primary) nucleic acid probe may be applied to a sample, e.g., simultaneously. For example, there may be at least 2, at least 5, at least 10, at least 25, at least 50, at least 75, at least 100, at least 300, at least 1,000, at least 3,000, at least 10,000, or at least 30,000 distinguishable nucleic acid probes that are applied to a sample, e.g., simultaneously or sequentially.

The target sequence may be positioned anywhere within the nucleic acid probe (or primary nucleic acid probe or encoding nucleic acid probe). The target sequence may contain

a region that is substantially complementary to a portion of a target nucleic acid. In some cases, the portions may be at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% complementary. In some cases, the target sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 50, at least 60, at least 65, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides in length. In some cases, the target sequence may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the target sequence may have a length of between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc. Typically, complementarity is determined on the basis of Watson-Crick nucleotide base pairing.

The target sequence of a (primary) nucleic acid probe may be determined with reference to a target nucleic acid suspected of being present within a cell or other sample. For example, a target nucleic acid to a protein may be determined using the protein's sequence, by determining the nucleic acids that are expressed to form the protein. In some cases, only a portion of the nucleic acids encoding the protein are used, e.g., having the lengths as discussed above. In addition, in some cases, more than one target sequence that can be used to identify a particular target may be used. For instance, multiple probes can be used, sequentially and/or simultaneously, that can bind to or hybridize to different regions of the same target. Hybridization typically refers to an annealing process by which complementary single-stranded nucleic acids associate through Watson-Crick nucleotide base pairing (e.g., hydrogen bonding, guanine-cytosine and adenine-thymine) to form double-stranded nucleic acid.

In some embodiments, a nucleic acid probe, such as a primary nucleic acid probe, may also comprise one or more "read" sequences. However, it should be understood that read sequences are not necessary in all cases. In some embodiments, the nucleic acid probe may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 or more, 20 or more, 32 or more, 40 or more, 50 or more, 64 or more, 75 or more, 100 or more, 128 or more read

sequences. The read sequences may be positioned anywhere within the nucleic acid probe. If more than one read sequence is present, the read sequences may be positioned next to each other, and/or interspersed with other sequences.

The read sequences, if present, may be of any length. If more than one read sequence is used, the read sequences may independently have the same or different lengths. For instance, the read sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 50, at least 60, at least 65, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides in length. In some cases, the read sequence may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the read sequence may have a length of between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc.

The read sequence may be arbitrary or random in some embodiments. In certain cases, the read sequences are chosen so as to reduce or minimize homology with other components of the cell or other sample, e.g., such that the read sequences do not themselves bind to or hybridize with other nucleic acids suspected of being within the cell or other sample. In some cases, the homology may be less than 10%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, or less than 1%. In some cases, there may be a homology of less than 20 basepairs, less than 18 basepairs, less than 15 basepairs, less than 14 basepairs, less than 13 basepairs, less than 12 basepairs, less than 11 basepairs, or less than 10 basepairs. In some cases, the basepairs are sequential.

In one set of embodiments, a population of nucleic acid probes may contain a certain number of read sequences, which may be less than the number of targets of the nucleic acid probes in some cases. Those of ordinary skill in the art will be aware that if there is one signaling entity and n read sequences, then in general 2^n-1 different nucleic acid targets may be uniquely identified. However, not all possible combinations need be used. For instance, a population of nucleic acid probes may target 12 different nucleic acid sequences, yet contain no more than 8 read sequences. As another example, a population of nucleic acids may target 140 different nucleic acid species, yet contain no more than 16 read sequences. Different

nucleic acid sequence targets may be separately identified by using different combinations of read sequences within each probe. For instance, each probe may contain 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, etc. or more read sequences. In some cases, a population of nucleic acid probes may each contain the same number of read sequences, although in other cases, there may be different numbers of read sequences present on the various probes.

As a non-limiting example, a first nucleic acid probe may contain a first target sequence, a first read sequence, and a second read sequence, while a second, different nucleic acid probe may contain a second target sequence, the same first read sequence, but a third read sequence instead of the second read sequence. Such probes may thereby be distinguished by determining the various read sequences present or associated with a given probe or location, as discussed herein.

In addition, the nucleic acid probes (and their corresponding, complimentary sites on the encoding probes), in certain embodiments, may be made using only 2 or only 3 of the 4 bases, such as leaving out all the "G"s or leaving out all of the "C"s within the probe. Sequences lacking either "G"s or "C"s may form very little secondary structure in certain embodiments, and can contribute to more uniform, faster hybridization.

In some embodiments, the nucleic acid probe may contain a signaling entity. It should be understood that signaling entities are not required in all cases, however; for instance, the nucleic acid probe may be determined using secondary nucleic acid probes in some embodiments, as is discussed in additional detail below. Examples of signaling entities that can be used are also discussed in more detail below.

Other components may also be present within a nucleic acid probe as well. For example, in one set of embodiments, one or more primer sequences may be present, e.g., to allow for enzymatic amplification of probes. Those of ordinary skill in the art will be aware of primer sequences suitable for applications such as amplification (e.g., using PCR or other suitable techniques). Many such primer sequences are available commercially. Other examples of sequences that may be present within a primary nucleic acid probe include, but are not limited to promoter sequences, operons, identification sequences, nonsense sequences, or the like.

Typically, a primer is a single-stranded or partially double-stranded nucleic acid (e.g., DNA) that serves as a starting point for nucleic acid synthesis, allowing polymerase enzymes such as nucleic acid polymerase to extend the primer and replicate the complementary strand. A primer is (e.g., is designed to be) complementary to and to hybridize to a target nucleic acid. In some embodiments, a primer is a synthetic primer. In some embodiments, a primer

is a non-naturally-occurring primer. A primer typically has a length of 10 to 50 nucleotides. For example, a primer may have a length of 10 to 40, 10 to 30, 10 to 20, 25 to 50, 15 to 40, 15 to 30, 20 to 50, 20 to 40, or 20 to 30 nucleotides. In some embodiments, a primer has a length of 18 to 24 nucleotides.

In addition, the components of the nucleic acid probe may be arranged in any suitable order. For instance, in one embodiment, the components may be arranged in a nucleic acid probe as: primer—read sequences—targeting sequence—read sequences—reverse primer. The "read sequences" in this structure may each contain any number (including 0) of read sequences, so long as at least one read sequence is present in the probe. Non-limiting example structures include primer—targeting sequence—read sequences—reverse primer, primer—read sequences—targeting sequence—reverse primer, targeting sequence—primer—targeting sequence—read sequences—reverse primer, targeting sequence—primer—read sequences—targeting sequence—reverse primer, primer—target sequence—read sequences—targeting sequence—reverse primer, targeting sequence—primer—read sequence—reverse primer, targeting sequence—read sequence—primer, read sequence—targeting sequence—primer, read sequence—primer—targeting sequence—reverse primer, etc. In addition, the reverse primer is optional in some embodiments, including in all of the above-described examples.

After introduction of the nucleic acid probes into a cell or other sample, the nucleic acid probes may be directly determined by determining signaling entities (if present), and/or the nucleic acid probes may be determined by using one or more secondary nucleic acid probes, in accordance with certain aspects of the invention. As mentioned, in some cases, the determination may be spatial, e.g., in two or three dimensions. In addition, in some cases, the determination may be quantitative, e.g., the amount or concentration of a primary nucleic acid probe (and of a target nucleic acid) may be determined. Additionally, the secondary probes may comprise any of a variety of entities able to hybridize a nucleic acid, e.g., DNA, RNA, LNA, and/or PNA, etc., depending on the application. Signaling entities are discussed in more detail below.

A secondary nucleic acid probe may contain a recognition sequence able to bind to or hybridize with a read sequence of a primary nucleic acid probe. In some cases, the binding is specific, or the binding may be such that a recognition sequence preferentially binds to or hybridizes with only one of the read sequences that are present. The secondary nucleic acid probe may also contain one or more signaling entities. If more than one secondary nucleic acid probe is used, the signaling entities may be the same or different.

The recognition sequences may be of any length, and multiple recognition sequences may be of the same or different lengths. If more than one recognition sequence is used, the recognition sequences may independently have the same or different lengths. For instance, the recognition sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, or at least 50 nucleotides in length. In some cases, the recognition sequence may be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the recognition sequence may have a length of between 10 and 30, between 20 and 40, or between 25 and 35 nucleotides, etc. In one embodiment, the recognition sequence is of the same length as the read sequence. In addition, in some cases, the recognition sequence may be at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 100% complementary to a read sequence of the primary nucleic acid probe.

As mentioned, in some cases, the secondary nucleic acid probe may comprise one or more signaling entities. Examples of signaling entities are discussed in more detail below.

As discussed, in certain aspects of the invention, nucleic acid probes are used that contain various "read sequences." For example, a population of primary nucleic acid probes may contain certain "read sequences" which can bind certain of the secondary nucleic acid probes, and the locations of the primary nucleic acid probes are determined within the sample using secondary nucleic acid probes, e.g., which comprise a signaling entity. As mentioned, in some cases, a population of read sequences may be combined in various combinations to produce different nucleic acid probes, e.g., such that a relatively small number of read sequences may be used to produce a relatively large number of different nucleic acid probes.

Thus, in some cases, a population of primary nucleic acid probes (or other nucleic acid probes) may each contain a certain number of read sequences, some of which are shared between different primary nucleic acid probes such that the total population of primary nucleic acid probes may contain a certain number of read sequences. A population of nucleic acid probes may have any suitable number of read sequences. For example, a population of primary nucleic acid probes may have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 etc. read sequences. More than 20 are also possible in some embodiments. In addition, in some cases, a population of nucleic acid probes may, in total, have 1 or more, 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or

more, 11 or more, 12 or more, 13 or more, 14 or more, 15 or more, 16 or more, 20 or more, 24 or more, 32 or more, 40 or more, 50 or more, 60 or more, 64 or more, 100 or more, 128 or more, etc. of possible read sequences present, although some or all of the probes may each contain more than one read sequence, as discussed herein. In addition, in some embodiments, the population of nucleic acid probes may have no more than 100, no more than 80, no more than 64, no more than 60, no more than 50, no more than 40, no more than 32, no more than 24, no more than 20, no more than 16, no more than 15, no more than 14, no more than 13, no more than 12, no more than 11, no more than 10, no more than 9, no more than 8, no more than 7, no more than 6, no more than 5, no more than 4, no more than 3, or no more than two read sequences present. Combinations of any of these are also possible, e.g., a population of nucleic acid probes may comprise between 10 and 15 read sequences in total.

As a non-limiting example of an approach to combinatorially producing a relatively large number of nucleic acid probes from a relatively small number of read sequences, in a population of 6 different types of nucleic acid probes, each comprising one or more read sequences, the total number of read sequences within the population may be no greater than 4. It should be understood that although 4 read sequences are used in this example for ease of explanation, in other embodiments, larger numbers of nucleic acid probes may be realized, for example, using 5, 8, 10, 16, 32, etc. or more read sequences, or any other suitable number of read sequences described herein, depending on the application. Referring now to Fig. 4A, if each of the primary nucleic acid probes contains two different read sequences, then by using 4 such read sequences (A, B, C, and D), up to 6 probes may be separately identified. It should be noted that in this example, the ordering of read sequences on a nucleic acid probe is not essential, i.e., "AB" and "BA" may be treated as being synonymous (although in other embodiments, the ordering of read sequences may be essential and "AB" and "BA" may not necessarily be synonymous). Similarly, if 5 read sequences are used (A, B, C, D, and E) in the population of primary nucleic acid probes, up to 10 probes may be separately identified, as is shown in Fig. 4B. For example, one of ordinary skill in the art would understand that, for k read sequences in a population with n read sequences on each probe, up to $\binom{n}{k}$ different probes may be produced, assuming that the ordering of read sequences is not essential; because not all of the probes need to have the same number of read sequences and not all combinations of read sequences need to be used in every embodiment, either more or less than this number of different probes may also be used in certain embodiments. In addition, it should also be understood that the number of read sequences on each probe need not be

identical in some embodiments. For instance example, some probes may contain 2 read sequences while other probes may contain 3 read sequences.

In some aspects, the read sequences and/or the pattern of binding of nucleic acid probes within a sample may be used to define an error-detecting and/or an error-correcting code, for example, to reduce or prevent misidentification or errors of the nucleic acids, e.g., as was discussed with reference to Fig. 3. Thus, for example, if binding is indicated (e.g., as determined using a signaling entity), then the location may be identified with a "1"; conversely, if no binding is indicated, then the location may be identified with a "0" (or vice versa, in some cases). Multiple rounds of binding determinations, e.g., using different nucleic acid probes, can then be used to create a "codeword," e.g., for that spatial location. In some embodiments, the codeword may be subjected to error detection and/or correction. For instance, the codewords may be organized such that, if no match is found for a given set of read sequences or binding pattern of nucleic acid probes, then the match may be identified as an error, and optionally, error correction may be applied sequences to determine the correct target for the nucleic acid probes. In some cases, the codewords may have fewer "letters" or positions that the total number of nucleic acids encoded by the codewords, e.g. where each codeword encodes a different nucleic acid.

Such error-detecting and/or the error-correction code may take a variety of forms. A variety of such codes have previously been developed in other contexts such as the telecommunications industry, such as Golay codes or Hamming codes. In one set of embodiments, the read sequences or binding patterns of the nucleic acid probes are assigned such that not every possible combination is assigned.

For example, if 4 read sequences are possible and a primary nucleic acid probe contains 2 read sequences, then up to 6 primary nucleic acid probes could be identified; but the number of primary nucleic acid probes used may be less than 6. Similarly, for k read sequences in a population with n read sequences on each primary nucleic acid probe, $\binom{n}{k}$ different probes may be produced, but the number of primary nucleic acid probes that are used may be any number more or less than $\binom{n}{k}$. In addition, these may be randomly assigned, or assigned in specific ways to increase the ability to detect and/or correct errors.

As another example, if multiple rounds of nucleic acid probes are used, the number of rounds may be arbitrarily chosen. If in each round, each target can give two possible outcomes, such as being detected or not being detected, up to 2^n different targets may be possible for n rounds of probes, but the number of nucleic acid targets that are actually used

may be any number less than 2^n. For example, if in each round, each target can give more than two possible outcomes, such as being detected in different color channels, more than 2^n (e.g. 3^n, 4^n ...) different targets may be possible for n rounds of probes. In some cases, the number of nucleic acid targets that are actually used may be any number less than this number. In addition, these may be randomly assigned, or assigned in specific ways to increase the ability to detect and/or correct errors.

For example, in one set of embodiments, the codewords or nucleic acid probes may be assigned within a code space such that the assignments are separated by a Hamming distance, which measures the number of incorrect "reads" in a given pattern that cause the nucleic acid probe to be misinterpreted as a different valid nucleic acid probe. In certain cases, the Hamming distance may be at least 2, at least 3, at least 4, at least 5, at least 6, or the like. In addition, in one set of embodiments, the assignments may be formed as a Hamming code, for instance, a Hamming(7, 4) code, a Hamming(15, 11) code, a Hamming(31, 26) code, a Hamming(63, 57) code, a Hamming(127, 120) code, etc. In another set of embodiments, the assignments may form a SECDED code, e.g., a SECDED(8,4) code, a SECDED(16,4) code, a SCEDED(16, 11) code, a SCEDED(22, 16) code, a SCEDED(39, 32) code, a SCEDED(72, 64) code, etc. In yet another set of embodiments, the assignments may form an extended binary Golay code, a perfect binary Golay code, or a ternary Golay code. In another set of embodiments, the assignments may represent a subset of the possible values taken from any of the codes described above.

For example, a code with the same error correcting properties of the SECDED code may be formed by using only binary words that contain a fixed number of '1' bits, such as 4, to encode the targets. In another set of embodiments, the assignments may represent a subset of the possible values taken from codes described above for the purpose of addressing asymmetric readout errors. For example, in some cases, a code in which the number of '1' bits may be fixed for all used binary words may eliminate the biased measurement of words with different numbers of '1's when the rate at which '0' bits are measured as '1's or '1' bits are measured as '0's are different.

Accordingly, in some embodiments, once the codeword is determined (e.g., as discussed herein), the codeword may be compared to the known nucleic acid codewords. If a match is found, then the nucleic acid target can be identified or determined. If no match is found, then an error in the reading of the codeword may be identified. In some cases, error correction can also be applied to determine the correct codeword, and thus resulting in the correct identity of the nucleic acid target. In some cases, the codewords may be selected such

that, assuming that there is only one error present, only one possible correct codeword is available, and thus, only one correct identity of the nucleic acid target is possible. In some cases, this may also be generalized to larger codeword spacings or Hamming distances; for instance, the codewords may be selected such that if two, three, or four errors are present (or more in some cases), only one possible correct codeword is available, and thus, only one correct identity of the nucleic acid targets is possible.

The error-correcting code may be a binary error-correcting code, or it may be based on other numbering systems, e.g., ternary or quaternary error-correcting codes. For instance, in one set of embodiments, more than one type of signaling entity may be used and assigned to different numbers within the error-correcting code. Thus, as a non-limiting example, a first signaling entity (or more than one signaling entity, in some cases) may be assigned as "1" and a second signaling entity (or more than one signaling entity, in some cases) may be assigned as "2" (with "0" indicating no signaling entity present), and the codewords distributed to define a ternary error-correcting code. Similarly, a third signaling entity may additionally be assigned as "3" to make a quaternary error-correcting code, etc.

As discussed above, in certain aspects, signaling entities are determined, e.g., to determine nucleic acid probes and/or to create codewords. In some cases, signaling entities within a sample may be determined, e.g., spatially, using a variety of techniques. In some embodiments, the signaling entities may be fluorescent, and techniques for determining fluorescence within a sample, such as fluorescence microscopy or confocal microscopy, may be used to spatially identify the positions of signaling entities within a cell. In some cases, the positions of entities within the sample may be determined in two or even three dimensions. In addition, in some embodiments, more than one signaling entity may be determined at a time (e.g., signaling entities with different colors or emissions), and/or sequentially.

In addition, in some embodiments, a confidence level for the identified nucleic acid target may be determined. For example, the confidence level may be determined using a ratio of the number of exact matches to the number of matches having one or more one-bit errors. In some cases, only matches having a confidence ratio greater than a certain value may be used. For instance, in certain embodiments, matches may be accepted only if the confidence ratio for the match is greater than about 0.01, greater than about 0.03, greater than about 0.05, greater than about 0.1, greater than about 0.3, greater than about 0.5, greater than about 1, greater than about 3, greater than about 5, greater than about 10, greater than about 30, greater than about 50, greater than about 100, greater than about 300, greater than about

500, greater than about 1000, or any other suitable value. In addition, in some embodiments, matches may be accepted only if the confidence ratio for the identified nucleic acid target is greater than an internal standard or false positive control by about 0.01, about 0.03, about 0.05, about 0.1, about 0.3, about 0.5, about 1, about 3, about 5, about 10, about 30, about 50, about 100, about 300, about 500, about 1000, or any other suitable value

In some embodiments, the spatial positions of the entities (and thus, nucleic acid probes that the entities may be associated with) may be determined at relatively high resolutions. For instance, the positions may be determined at spatial resolutions of better than about 100 micrometers, better than about 30 micrometers, better than about 10 micrometers, better than about 3 micrometers, better than about 1 micrometer, better than about 800 nm, better than about 600 nm, better than about 500 nm, better than about 400 nm, better than about 300 nm, better than about 200 nm, better than about 100 nm, better than about 90 nm, better than about 80 nm, better than about 70 nm, better than about 60 nm, better than about 50 nm, better than about 40 nm, better than about 30 nm, better than about 20 nm, or better than about 10 nm, etc.

There are a variety of techniques able to determine or image the spatial positions of entities optically, e.g., using fluorescence microscopy. In some cases, the spatial positions may be determined at super resolutions, or at resolutions better than the wavelength of light or the diffraction limit. Non-limiting examples include STORM (stochastic optical reconstruction microscopy), STED (stimulated emission depletion microscopy), NSOM (Near-field Scanning Optical Microscopy), 4Pi microscopy, SIM (Structured Illumination Microscopy), SMI (Spatially Modulated Illumination) microscopy, RESOLFT (Reversible Saturable Optically Linear Fluorescence Transition Microscopy), GSD (Ground State Depletion Microscopy), SSIM (Saturated Structured-Illumination Microscopy), SPDM (Spectral Precision Distance Microscopy), Photo-Activated Localization Microscopy (PALM), Fluorescence Photoactivation Localization Microscopy (FPALM), LIMON (3D Light Microscopical Nanosizing Microscopy), Super-resolution optical fluctuation imaging (SOFI), or the like. See, e.g., U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*; or Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*, each incorporated herein by reference in their entireties.

As an illustrative non-limiting example, in one set of embodiments, the sample may be imaged with a high numerical aperture, oil immersion objective with 100X magnification and light collected on an electron-multiplying CCD camera. In another example, the sample could be imaged with a high numerical aperture, oil immersion lens with 40X magnification and light collected with a wide-field scientific CMOS camera. With different combinations of objectives and cameras, a single field of view may correspond to no less than 40 x 40 microns, 80 x 80 microns, 120 x 120 microns, 240 x 240 microns, 340 x 340 microns, or 500 x 500 microns, etc. in various non-limiting embodiments. Similarly, a single camera pixel may correspond, in some embodiments, to regions of the sample of no less than 80x80 nm, 120x120 nm, 160x160 nm, 240x240 nm, or 300x300 nm, etc. In another example, the sample may be imaged with a low numerical aperture, air lens with 10X magnification and light collected with a sCMOS camera. In additional embodiments, the sample may be optically sectioned by illuminating it via a single or multiple scanned diffraction limited foci generated either by scanning mirrors or a spinning disk and the collected passed through a single or multiple pinholes. In another embodiment, the sample may also be illuminated via thin sheet of light generated via any one of multiple methods known to those versed in the art.

In one embodiment, the sample may be illuminated by single Gaussian mode laser lines. In some embodiments, the illumination profiled may be flattened by passing these laser lines through a multimode fiber that is vibrated via piezo-electric or other mechanical means. In some embodiments, the illumination profile may be flattened by passing single-mode, Gaussian beams through a variety of refractive beam shapers, such as the piShaper or a series of stacked Powell lenses. In yet another set of embodiments, the Gaussian beams may be passed through a variety of different diffusing elements, such as ground glass or engineered diffusers, which may be spun in some cases at high speeds to remove residual laser speckle. In yet another embodiment, laser illumination may be passed through a series of lenslet arrays to produce overlapping images of the illumination that approximate a flat illumination field.

In some embodiments, the centroids of the spatial positions of the entities may be determined. For example, a centroid of a signaling entity may be determined within an image or series of images using image analysis algorithms known to those of ordinary skill in the art. In some cases, the algorithms may be selected to determine non-overlapping single emitters and/or partially overlapping single emitters in a sample. Non-limiting examples of suitable techniques include a maximum likelihood algorithm, a least squares algorithm, a

Bayesian algorithm, a compressed sensing algorithm, or the like. Combinations of these techniques may also be used in some cases.

In addition, the signaling entity may be inactivated in some cases. For example, in some embodiments, a first secondary nucleic acid probe containing a signaling entity may be applied to a sample that can recognize a first read sequence, then the first secondary nucleic acid probe can be inactivated before a second secondary nucleic acid probe is applied to the sample. If multiple signaling entities are used, the same or different techniques may be used to inactivate the signaling entities, and some or all of the multiple signaling entities may be inactivated, e.g., sequentially or simultaneously.

Inactivation may be caused by removal of the signaling entity (e.g., from the sample, or from the nucleic acid probe, etc.), and/or by chemically altering the signaling entity in some fashion, e.g., by photobleaching the signaling entity, bleaching or chemically altering the structure of the signaling entity, e.g., by reduction, etc.). For instance, in one set of embodiments, a fluorescent signaling entity may be inactivated by chemical or optical techniques such as oxidation, photobleaching, chemically bleaching, stringent washing or enzymatic digestion or reaction by exposure to an enzyme, dissociating the signaling entity from other components (e.g., a probe), chemical reaction of the signaling entity (e.g., to a reactant able to alter the structure of the signaling entity) or the like. For instance, bleaching may occur by exposure to oxygen, reducing agents, or the signaling entity could be chemically cleaved from the nucleic acid probe and washed away via fluid flow.

In some embodiments, various nucleic acid probes (including primary and/or secondary nucleic acid probes) may include one or more signaling entities. If more than one nucleic acid probe is used, the signaling entities may each by the same or different. In certain embodiments, a signaling entity is any entity able to emit light. For instance, in one embodiment, the signaling entity is fluorescent. In other embodiments, the signaling entity may be phosphorescent, radioactive, absorptive, etc. In some cases, the signaling entity is any entity that can be determined within a sample at relatively high resolutions, e.g., at resolutions better than the wavelength of visible light or the diffraction limit. The signaling entity may be, for example, a dye, a small molecule, a peptide or protein, or the like. The signaling entity may be a single molecule in some cases. If multiple secondary nucleic acid probes are used, the nucleic acid probes may comprise the same or different signaling entities.

Non-limiting examples of signaling entities include fluorescent entities (fluorophores) or phosphorescent entities, for example, cyanine dyes (e.g., Cy2, Cy3, Cy3B, Cy5, Cy5.5,

Cy7, etc.), Alexa Fluor dyes, Atto dyes, photoswtichable dyes, photoactivatable dyes, fluorescent dyes, metal nanoparticles, semiconductor nanoparticles or "quantum dots", fluorescent proteins such as GFP (Green Fluorescent Protein), or photoactivabale fluorescent proteins, such as PAGFP, PSCFP, PSCFP2, Dendra, Dendra2, EosFP, tdEos, mEos2, mEos3, PAmCherry, PAtagRFP, mMaple, mMaple2, and mMaple3. Other suitable signaling entities are known to those of ordinary skill in the art. See, e.g., U.S. Pat. No. 7,838,302 or U.S. Pat. Apl. Ser. No. 61/979,436, each incorporated herein by reference in its entirety.

In one set of embodiments, the signaling entity may be attached to an oligonucleotide sequence via a bond that can be cleaved to release the signaling entity. In one set of embodiments, a fluorophore may be conjugated to an oligonucleotide via a cleavable bond, such as a photocleavable bond. Non-limiting examples of photocleavable bonds include, but are not limited to, 1-(2-nitrophenyl)ethyl, 2-nitrobenzyl, biotin phosphoramidite, acrylic phosphoramidite, diethylaminocoumarin, 1-(4,5-dimethoxy-2-nitrophenyl)ethyl, cyclo-dodecyl (dimethoxy-2-nitrophenyl)ethyl, 4-aminomethyl-3-nitrobenzyl, (4-nitro-3-(1-chlorocarbonyloxyethyl)phenyl)methyl-S-acetylthioic acid ester, (4-nitro-3-(1-thlorocarbonyloxyethyl)phenyl)methyl-3-(2-pyridyldithiopropionic acid) ester, 3-(4,4'-dimethoxytrityl)-1-(2-nitrophenyl)-propane-1,3-diol-[2- cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-trifluoroacetylcaproamidomethyl)phenyl]-ethyl-[2-cyano-ethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-(4,4'-dimethoxytrityloxy)butyramidomethyl)phenyl]-ethyl-[2-cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-(N-(4,4'-dimethoxytrityl))-biotinamidocaproamido-methyl)phenyl]-ethyl-[2-cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, or similar linkers. In another set of embodiments, the fluorophore may be conjugated to an oligonucleotide via a disulfide bond. The disulfide bond may be cleaved by a variety of reducing agents such as, but not limited to, dithiothreitol, dithioerythritol, beta-mercaptoethanol, sodium borohydride, thioredoxin, glutaredoxin, trypsinogen, hydrazine, diisobutylaluminum hydride, oxalic acid, formic acid, ascorbic acid, phosphorous acid, tin chloride, glutathione, thioglycolate, 2,3-dimercaptopropanol, 2-mercaptoethylamine, 2-aminoethanol, tris(2-carboxyethyl)phosphine, bis(2-mercaptoethyl) sulfone, N,N'-dimethyl-N,N'-bis(mercaptoacetyl)hydrazine, 3-mercaptoproptionate, dimethylformamide, thiopropyl-agarose, tri-*n*-butylphosphine, cysteine, iron sulfate, sodium sulfite, phosphite, hypophosphite, phosphorothioate, or the like, and/or combinations of any of these. In another embodiment, the fluorophore may be conjugated to an oligonucleotide via one or more phosphorothioate modified nucleotides in which the sulfur modification replaces the bridging and/or non-bridging oxygen. The fluorophore may be

cleaved from the oligonucleotide, in certain embodiments, via addition of compounds such as but not limited to iodoethanol, iodine mixed in ethanol, silver nitrate, or mercury chloride. In yet another set of embodiments, the signaling entity may be chemically inactivated through reduction or oxidation. For example, in one embodiment, a chromophore such as Cy5 or Cy7 may be reduced using sodium borohydride to a stable, non-fluorescence state. In still another set of embodiments, a fluorophore may be conjugated to an oligonucleotide via an azo bond, and the azo bond may be cleaved with 2-[(2-N-arylamino)phenylazo]pyridine. In yet another set of embodiments, a fluorophore may be conjugated to an oligonucleotide via a suitable nucleic acid segment that can be cleaved upon suitable exposure to DNAse, e.g., an exodeoxyribonuclease or an endodeoxyribonuclease. Examples include, but are not limited to, deoxyribonuclease I or deoxyribonuclease II. In one set of embodiments, the cleavage may occur via a restriction endonuclease. Non-limiting examples of potentially suitable restriction endonucleases include BamHI, BsrI, NotI, XmaI, PspAI, DpnI, MboI, MnlI, Eco57I, Ksp632I, DraIII, AhaII, SmaI, MluI, HpaI, ApaI, BclI, BstEII, TaqI, EcoRI, SacI, HindII, HaeII, DraII, Tsp509I, Sau3AI, PacI, etc. Over 3000 restriction enzymes have been studied in detail, and more than 600 of these are available commercially. In yet another set of embodiments, a fluorophore may be conjugated to biotin, and the oligonucleotide conjugated to avidin or streptavidin. An interaction between biotin and avidin or streptavidin allows the fluorophore to be conjugated to the oligonucleotide, while sufficient exposure to an excess of addition, free biotin could "outcompete" the linkage and thereby cause cleavage to occur. In addition, in another set of embodiments, the probes may be removed using corresponding "toe-hold-probes," which comprise the same sequence as the probe, as well as an extra number of bases of homology to the encoding probes (e.g., 1-20 extra bases, for example, 5 extra bases). These probes may remove the labeled readout probe through a strand-displacement interaction.

As used herein, the term "light" generally refers to electromagnetic radiation, having any suitable wavelength (or equivalently, frequency). For instance, in some embodiments, the light may include wavelengths in the optical or visual range (for example, having a wavelength of between about 400 nm and about 700 nm, i.e., "visible light"), infrared wavelengths (for example, having a wavelength of between about 300 micrometers and 700 nm), ultraviolet wavelengths (for example, having a wavelength of between about 400 nm and about 10 nm), or the like. In certain cases, as discussed in detail below, more than one entity may be used, i.e., entities that are chemically different or distinct, for example,

structurally. However, in other cases, the entities may be chemically identical or at least substantially chemically identical.

In one set of embodiments, the signaling entity is "switchable," i.e., the entity can be switched between two or more states, at least one of which emits light having a desired wavelength. In the other state(s), the entity may emit no light, or emit light at a different wavelength. For instance, an entity may be "activated" to a first state able to produce light having a desired wavelength, and "deactivated" to a second state not able to emit light of the same wavelength. An entity is "photoactivatable" if it can be activated by incident light of a suitable wavelength. As a non-limiting example, Cy5, can be switched between a fluorescent and a dark state in a controlled and reversible manner by light of different wavelengths, i.e., 633 nm (or 642nm, 647nm, 656 nm) red light can switch or deactivate Cy5 to a stable dark state, while 405 nm green light can switch or activate the Cy5 back to the fluorescent state. In some cases, the entity can be reversibly switched between the two or more states, e.g., upon exposure to the proper stimuli. For example, a first stimuli (e.g., a first wavelength of light) may be used to activate the switchable entity, while a second stimuli (e.g., a second wavelength of light) may be used to deactivate the switchable entity, for instance, to a non-emitting state. Any suitable method may be used to activate the entity. For example, in one embodiment, incident light of a suitable wavelength may be used to activate the entity to emit light, i.e., the entity is "photoswitchable." Thus, the photoswitchable entity can be switched between different light-emitting or non-emitting states by incident light, e.g., of different wavelengths. The light may be monochromatic (e.g., produced using a laser) or polychromatic. In another embodiment, the entity may be activated upon stimulation by electric field and/or magnetic field. In other embodiments, the entity may be activated upon exposure to a suitable chemical environment, e.g., by adjusting the pH, or inducing a reversible chemical reaction involving the entity, etc. Similarly, any suitable method may be used to deactivate the entity, and the methods of activating and deactivating the entity need not be the same. For instance, the entity may be deactivated upon exposure to incident light of a suitable wavelength, or the entity may be deactivated by waiting a sufficient time.

Typically, a "switchable" entity can be identified by one of ordinary skill in the art by determining conditions under which an entity in a first state can emit light when exposed to an excitation wavelength, switching the entity from the first state to the second state, e.g., upon exposure to light of a switching wavelength, then showing that the entity, while in the second state can no longer emit light (or emits light at a much reduced intensity) when exposed to the excitation wavelength.

In one set of embodiments, as discussed, a switchable entity may be switched upon exposure to light. In some cases, the light used to activate the switchable entity may come from an external source, e.g., a light source such as a laser light source, another light-emitting entity proximate the switchable entity, etc. The second, light emitting entity, in some cases, may be a fluorescent entity, and in certain embodiments, the second, light-emitting entity may itself also be a switchable entity.

In some embodiments, the switchable entity includes a first, light-emitting portion (e.g., a fluorophore), and a second portion that activates or "switches" the first portion. For example, upon exposure to light, the second portion of the switchable entity may activate the first portion, causing the first portion to emit light. Examples of activator portions include, but are not limited to, Alexa Fluor 405 (Invitrogen), Alexa Fluor 488 (Invitrogen), Cy2 (GE Healthcare), Cy3 (GE Healthcare), Cy3B (GE Healthcare), Cy3.5 (GE Healthcare), or other suitable dyes. Examples of light-emitting portions include, but are not limited to, Cy5, Cy5.5 (GE Healthcare), Cy7 (GE Healthcare), Alexa Fluor 647 (Invitrogen), Alexa Fluor 680 (Invitrogen), Alexa Fluor 700 (Invitrogen), Alexa Fluor 750 (Invitrogen), Alexa Fluor 790 (Invitrogen), DiD, DiR, YOYO-3 (Invitrogen), YO-PRO-3 (Invitrogen), TOT-3 (Invitrogen), TO-PRO-3 (Invitrogen) or other suitable dyes. These may linked together, e.g., covalently, for example, directly, or through a linker, e.g., forming compounds such as, but not limited to, Cy5-Alexa Fluor 405, Cy5-Alexa Fluor 488, Cy5-Cy2, Cy5-Cy3, Cy5-Cy3.5, Cy5.5-Alexa Fluor 405, Cy5.5-Alexa Fluor 488, Cy5.5-Cy2, Cy5.5-Cy3, Cy5.5-Cy3.5, Cy7-Alexa Fluor 405, Cy7-Alexa Fluor 488, Cy7-Cy2, Cy7-Cy3, Cy7-Cy3.5, Alexa Fluor 647-Alexa Fluor 405, Alexa Fluor 647-Alexa Fluor 488, Alexa Fluor 647-Cy2, Alexa Fluor 647-Cy3, Alexa Fluor 647-Cy3.5, Alexa Fluor 750-Alexa Fluor 405, Alexa Fluor 750-Alexa Fluor 488, Alexa Fluor 750-Cy2, Alexa Fluor 750-Cy3, or Alexa Fluor 750-Cy3.5. Those of ordinary skill in the art will be aware of the structures of these and other compounds, many of which are available commercially. The portions may be linked via a covalent bond, or by a linker, such as those described in detail below. Other light-emitting or activator portions may include portions having two quaternized nitrogen atoms joined by a polymethine chain, where each nitrogen is independently part of a heteroaromatic moiety, such as pyrrole, imidazole, thiazole, pyridine, quinoine, indole, benzothiazole, etc., or part of a nonaromatic amine. In some cases, there may be 5, 6, 7, 8, 9, or more carbon atoms between the two nitrogen atoms.

In certain cases, the light-emitting portion and the activator portions, when isolated from each other, may each be fluorophores, i.e., entities that can emit light of a certain,

emission wavelength when exposed to a stimulus, for example, an excitation wavelength. However, when a switchable entity is formed that comprises the first fluorophore and the second fluorophore, the first fluorophore forms a first, light-emitting portion and the second fluorophore forms an activator portion that switches that activates or "switches" the first portion in response to a stimulus. For example, the switchable entity may comprise a first fluorophore directly bonded to the second fluorophore, or the first and second entity may be connected via a linker or a common entity. Whether a pair of light-emitting portion and activator portion produces a suitable switchable entity can be tested by methods known to those of ordinary skills in the art. For example, light of various wavelength can be used to stimulate the pair and emission light from the light-emitting portion can be measured to determined wither the pair makes a suitable switch.

As a non-limiting example, Cy3 and Cy5 may be linked together to form such an entity. In this example, Cy3 is an activator portion that is able to activate Cy5, the light-emission portion. Thus, light at or near the absorption maximum (e.g., near 532 nm light for Cy3) of the activation or second portion of the entity may cause that portion to activate the first, light-emitting portion, thereby causing the first portion to emit light (e.g., near 647 nm for Cy5). See, e.g., U.S. Pat. No. 7,838,302, incorporated herein by reference in its entirety. In some cases, the first, light-emitting portion can subsequently be deactivated by any suitable technique (e.g., by directing 647 nm red light to the Cy5 portion of the molecule).

Other non-limiting examples of potentially suitable activator portions include 1,5 IAEDANS, 1,8-ANS, 4-Methylumbelliferone, 5-carboxy-2,7-dichlorofluorescein, 5-Carboxyfluorescein (5-FAM), 5-Carboxynapthofluorescein, 5-Carboxytetramethylrhodamine (5-TAMRA), 5-FAM (5-Carboxyfluorescein), 5-HAT (Hydroxy Tryptamine), 5-Hydroxy Tryptamine (HAT), 5-ROX (carboxy-X-rhodamine), 5-TAMRA (5-Carboxytetramethylrhodamine), 6-Carboxyrhodamine 6G, 6-CR 6G, 6-JOE, 7-Amino-4-methylcoumarin, 7-Aminoactinomycin D (7-AAD), 7-Hydroxy-4-methylcoumarin, 9-Amino-6-chloro-2-methoxyacridine, ABQ, Acid Fuchsin, ACMA (9-Amino-6-chloro-2-methoxyacridine), Acridine Orange, Acridine Red, Acridine Yellow, Acriflavin, Acriflavin Feulgen SITSA, Alexa Fluor 350, Alexa Fluor 405, Alexa Fluor 430, Alexa Fluor 488, Alexa Fluor 500, Alexa Fluor 514, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 555, Alexa Fluor 568, Alexa Fluor 594, Alexa Fluor 610, Alexa Fluor 633, Alexa Fluor 635, Alizarin Complexon, Alizarin Red, AMC, AMCA-S, AMCA (Aminomethylcoumarin), AMCA-X, Aminoactinomycin D, Aminocoumarin, Aminomethylcoumarin (AMCA), Anilin Blue, Anthrocyl stearate, APTRA-BTC, APTS, Astrazon Brilliant Red 4G, Astrazon Orange R,

Astrazon Red 6B, Astrazon Yellow 7 GLL, Atabrine, ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 520, ATTO 532, ATTO 550, ATTO 565, ATTO 590, ATTO 594, ATTO 610, ATTO 611X, ATTO 620, ATTO 633, ATTO 635, ATTO 647, ATTO 647N, ATTO 655, ATTO 680, ATTO 700, ATTO 725, ATTO 740, ATTO-TAG CBQCA, ATTO-TAG FQ, Auramine, Aurophosphine G, Aurophosphine, BAO 9 (Bisaminophenyloxadiazole), BCECF (high pH), BCECF (low pH), Berberine Sulphate, Bimane, Bisbenzamide, Bisbenzimide (Hoechst), bis-BTC, Blancophor FFG, Blancophor SV, BOBO -1, BOBO -3, Bodipy 492/515, Bodipy 493/503, Bodipy 500/510, Bodipy 505/515, Bodipy 530/550, Bodipy 542/563, Bodipy 558/568, Bodipy 564/570, Bodipy 576/589, Bodipy 581/591, Bodipy 630/650-X, Bodipy 650/665-X, Bodipy 665/676, Bodipy Fl, Bodipy FL ATP, Bodipy Fl-Ceramide, Bodipy R6G, Bodipy TMR, Bodipy TMR-X conjugate, Bodipy TMR-X, SE, Bodipy TR, Bodipy TR ATP, Bodipy TR-X SE, BO-PRO -1, BO-PRO -3, Brilliant Sulphoflavin FF, BTC, BTC-5N, Calcein, Calcein Blue, Calcium Crimson, Calcium Green, Calcium Green-1 Ca^{2+} Dye, Calcium Green-2 Ca^{2+}, Calcium Green-5N Ca^{2+}, Calcium Green-C18 Ca^{2+}, Calcium Orange, Calcofluor White, Carboxy-X-rhodamine (5-ROX), Cascade Blue, Cascade Yellow, Catecholamine, CCF2 (GeneBlazer), CFDA, Chromomycin A, Chromomycin A, CL-NERF, CMFDA, Coumarin Phalloidin, CPM Methylcoumarin, CTC, CTC Formazan, Cy2, Cy3.1 8, Cy3.5, Cy3, Cy5.1 8, cyclic AMP Fluorosensor (FiCRhR), Dabcyl, Dansyl, Dansyl Amine, Dansyl Cadaverine, Dansyl Chloride, Dansyl DHPE, Dansyl fluoride, DAPI, Dapoxyl, Dapoxyl 2, Dapoxyl 3' DCFDA, DCFH (Dichlorodihydrofluorescein Diacetate), DDAO, DHR (Dihydorhodamine 123), Di-4-ANEPPS, Di-8-ANEPPS (non-ratio), DiA (4-Di-16-ASP), Dichlorodihydrofluorescein Diacetate (DCFH), DiD - Lipophilic Tracer, DiD (DiIC18(5)), DIDS, Dihydorhodamine 123 (DHR), DiI (DiIC18(3)), Dinitrophenol, DiO (DiOC18(3)), DiR, DiR (DiIC18(7)), DM-NERF (high pH), DNP, Dopamine, DTAF, DY-630-NHS, DY-635-NHS, DyLight 405, DyLight 488, DyLight 549, DyLight 633, DyLight 649, DyLight 680, DyLight 800, ELF 97, Eosin, Erythrosin, Erythrosin ITC, Ethidium Bromide, Ethidium homodimer -1 (EthD-1), Euchrysin, EukoLight, Europium (III) chloride, Fast Blue, FDA, Feulgen (Pararosaniline), FIF (Formaldehyd Induced Fluorescence), FITC, Flazo Orange, Fluo-3, Fluo-4, Fluorescein (FITC), Fluorescein Diacetate, Fluoro-Emerald, Fluoro-Gold (Hydroxystilbamidine), Fluor-Ruby, FluorX, FM 1-43, FM 4-46, Fura Red (high pH), Fura Red/Fluo-3, Fura-2, Fura-2/BCECF, Genacryl Brilliant Red B, Genacryl Brilliant Yellow 10GF, Genacryl Pink 3G, Genacryl Yellow 5GF, GeneBlazer (CCF2), Gloxalic Acid, Granular blue, Haematoporphyrin, Hoechst 33258, Hoechst 33342, Hoechst 34580, HPTS,

Hydroxycoumarin, Hydroxystilbamidine (FluoroGold), Hydroxytryptamine, Indo-1, high calcium, Indo-1, low calcium, Indodicarbocyanine (DiD), Indotricarbocyanine (DiR), Intrawhite Cf, JC-1, JO-JO-1, JO-PRO-1, LaserPro, Laurodan, LDS 751 (DNA), LDS 751 (RNA), Leucophor PAF, Leucophor SF, Leucophor WS, Lissamine Rhodamine, Lissamine Rhodamine B, Calcein/Ethidium homodimer, LOLO-1, LO-PRO-1, Lucifer Yellow, Lyso Tracker Blue, Lyso Tracker Blue-White, Lyso Tracker Green, Lyso Tracker Red, Lyso Tracker Yellow, LysoSensor Blue, LysoSensor Green, LysoSensor Yellow/Blue, Mag Green, Magdala Red (Phloxin B), Mag-Fura Red, Mag-Fura-2, Mag-Fura-5, Mag-Indo-1, Magnesium Green, Magnesium Orange, Malachite Green, Marina Blue, Maxilon Brilliant Flavin 10 GFF, Maxilon Brilliant Flavin 8 GFF, Merocyanin, Methoxycoumarin, Mitotracker Green FM, Mitotracker Orange, Mitotracker Red, Mitramycin, Monobromobimane, Monobromobimane (mBBr-GSH), Monochlorobimane, MPS (Methyl Green Pyronine Stilbene), NBD, NBD Amine, Nile Red, Nitrobenzoxadidole, Noradrenaline, Nuclear Fast Red, Nuclear Yellow, Nylosan Brilliant Iavin E8G, Oregon Green, Oregon Green 488-X, Oregon Green, Oregon Green 488, Oregon Green 500, Oregon Green 514, Pacific Blue, Pararosaniline (Feulgen), PBFI, Phloxin B (Magdala Red), Phorwite AR, Phorwite BKL, Phorwite Rev, Phorwite RPA, Phosphine 3R, PKH26 (Sigma), PKH67, PMIA, Pontochrome Blue Black, POPO-1, POPO-3, PO-PRO-1, PO-PRO-3, Primuline, Procion Yellow, Propidium Iodid (PI), PyMPO, Pyrene, Pyronine, Pyronine B, Pyrozal Brilliant Flavin 7GF, QSY 7, Quinacrine Mustard, Resorufin, RH 414, Rhod-2, Rhodamine, Rhodamine 110, Rhodamine 123, Rhodamine 5 GLD, Rhodamine 6G, Rhodamine B, Rhodamine B 200, Rhodamine B extra, Rhodamine BB, Rhodamine BG, Rhodamine Green, Rhodamine Phallicidine, Rhodamine Phalloidine, Rhodamine Red, Rhodamine WT, Rose Bengal, S65A, S65C, S65L, S65T, SBFI, Serotonin, Sevron Brilliant Red 2B, Sevron Brilliant Red 4G, Sevron Brilliant Red B, Sevron Orange, Sevron Yellow L, SITS, SITS (Primuline), SITS (Stilbene Isothiosulphonic Acid), SNAFL calcein, SNAFL-1, SNAFL-2, SNARF calcein, SNARF1, Sodium Green, SpectrumAqua, SpectrumGreen, SpectrumOrange, Spectrum Red, SPQ (6-methoxy-N-(3-sulfopropyl)quinolinium), Stilbene, Sulphorhodamine B can C, Sulphorhodamine Extra, SYTO 11, SYTO 12, SYTO 13, SYTO 14, SYTO 15, SYTO 16, SYTO 17, SYTO 18, SYTO 20, SYTO 21, SYTO 22, SYTO 23, SYTO 24, SYTO 25, SYTO 40, SYTO 41, SYTO 42, SYTO 43, SYTO 44, SYTO 45, SYTO 59, SYTO 60, SYTO 61, SYTO 62, SYTO 63, SYTO 64, SYTO 80, SYTO 81, SYTO 82, SYTO 83, SYTO 84, SYTO 85, SYTOX Blue, SYTOX Green, SYTOX Orange, Tetracycline, Tetramethylrhodamine (TAMRA), Texas Red, Texas Red-X conjugate, Thiadicarbocyanine (DiSC3), Thiazine Red

R, Thiazole Orange, Thioflavin 5, Thioflavin S, Thioflavin TCN, Thiolyte, Thiozole Orange, Tinopol CBS (Calcofluor White), TMR, TO-PRO-1, TO-PRO-3, TO-PRO-5, TOTO-1, TOTO-3, TRITC (tetramethylrodamine isothiocyanate), True Blue, TruRed, Ultralite, Uranine B, Uvitex SFC, WW 781, X-Rhodamine, XRITC, Xylene Orange, Y66F, Y66H, Y66W, YO-PRO-1, YO-PRO-3, YOYO-1, YOYO-3, SYBR Green, Thiazole orange (interchelating dyes), or combinations thereof.

Another aspect of the invention is directed to a computer-implemented method. For instance, a computer and/or an automated system may be provided that is able to automatically and/or repetitively perform any of the methods described herein. As used herein, "automated" devices refer to devices that are able to operate without human direction, i.e., an automated device can perform a function during a period of time after any human has finished taking any action to promote the function, e.g. by entering instructions into a computer to start the process. Typically, automated equipment can perform repetitive functions after this point in time. The processing steps may also be recorded onto a machine-readable medium in some cases.

For example, in some cases, a computer may be used to control imaging of the sample, e.g., using fluorescence microscopy, STORM or other super-resolution techniques such as those described herein. In some cases, the computer may also control operations such as drift correction, physical registration, hybridization and cluster alignment in image analysis, cluster decoding (e.g., fluorescent cluster decoding), error detection or correction (e.g., as discussed herein), noise reduction, identification of foreground features from background features (such as noise or debris in images), or the like. As an example, the computer may be used to control activation and/or excitation of signaling entities within the sample, and/or the acquisition of images of the signaling entities. In one set of embodiments, a sample may be excited using light having various wavelengths and/or intensities, and the sequence of the wavelengths of light used to excite the sample may be correlated, using a computer, to the images acquired of the sample containing the signaling entities. For instance, the computer may apply light having various wavelengths and/or intensities to a sample to yield different average numbers of signaling entities in each region of interest (e.g., one activated entity per location, two activated entities per location, etc.). In some cases, this information may be used to construct an image and/or determine the locations of the signaling entities, in some cases at high resolutions, as noted above.

In some aspects, the sample is positioned on a microscope. In some cases, the microscope may contain one or more channels, such as microfluidic channels, to direct or

control fluid to or from the sample. For instance, in one embodiment, nucleic acid probes such as those discussed herein may be introduced and/or removed from the sample by flowing fluid through one or more channels to or from the sample. In some cases, there may also be one or more chambers or reservoirs for holding fluid, e.g., in fluidic communication with the channel, and/or with the sample. Those of ordinary skill in the art will be familiar with channels, including microfluidic channels, for moving fluid to or from a sample.

As used herein, "microfluidic," "microscopic," "microscale," the "micro-" prefix (for example, as in "microchannel"), and the like generally refers to elements or articles having widths or diameters of less than about 1 mm, and less than about 100 microns (micrometers) in some cases. In some embodiments, larger channels may be used instead of, or in conjunction with, microfluidic channels for any of the embodiments discussed herein. For examples, channels having widths or diameters of less than about 10 mm, less than about 9 mm, less than about 8 mm, less than about 7 mm, less than about 6 mm, less than about 5 mm, less than about 4 mm, less than about 3 mm, or less than about 2 mm may be used in certain instances. In some cases, the element or article includes a channel through which a fluid can flow. In all embodiments, specified widths can be a smallest width (i.e. a width as specified where, at that location, the article can have a larger width in a different dimension), or a largest width (i.e. where, at that location, the article has a width that is no wider than as specified, but can have a length that is greater). Thus, for instance, the microfluidic channel may have an average cross-sectional dimension (e.g., perpendicular to the direction of flow of fluid in the microfluidic channel) of less than about 1 mm, less than about 500 microns, less than about 300 microns, or less than about 100 microns. In some cases, the microfluidic channel may have an average diameter of less than about 60 microns, less than about 50 microns, less than about 40 microns, less than about 30 microns, less than about 25 microns, less than about 10 microns, less than about 5 microns, less than about 3 microns, or less than about 1 micron.

A "channel," as used herein, means a feature on or in an article (e.g., a substrate) that at least partially directs the flow of a fluid. In some cases, the channel may be formed, at least in part, by a single component, e.g. an etched substrate or molded unit. The channel can have any cross-sectional shape, for example, circular, oval, triangular, irregular, square or rectangular (having any aspect ratio), or the like, and can be covered or uncovered (i.e., open to the external environment surrounding the channel). In embodiments where the channel is completely covered, at least one portion of the channel can have a cross-section that is

completely enclosed, and/or the entire channel may be completely enclosed along its entire length with the exception of its inlet and outlet.

A channel may have any aspect ratio, e.g., an aspect ratio (length to average cross-sectional dimension) of at least about 2:1, more typically at least about 3:1, at least about 5:1, at least about 10:1, etc. As used herein, a "cross-sectional dimension," in reference to a fluidic or microfluidic channel, is measured in a direction generally perpendicular to fluid flow within the channel. A channel generally will include characteristics that facilitate control over fluid transport, e.g., structural characteristics and/or physical or chemical characteristics (hydrophobicity vs. hydrophilicity) and/or other characteristics that can exert a force (e.g., a containing force) on a fluid. The fluid within the channel may partially or completely fill the channel. In some cases the fluid may be held or confined within the channel or a portion of the channel in some fashion, for example, using surface tension (e.g., such that the fluid is held within the channel within a meniscus, such as a concave or convex meniscus). In an article or substrate, some (or all) of the channels may be of a particular size or less, for example, having a largest dimension perpendicular to fluid flow of less than about 5 mm, less than about 2 mm, less than about 1 mm, less than about 500 microns, less than about 200 microns, less than about 100 microns, less than about 60 microns, less than about 50 microns, less than about 40 microns, less than about 30 microns, less than about 25 microns, less than about 10 microns, less than about 3 microns, less than about 1 micron, less than about 300 nm, less than about 100 nm, less than about 30 nm, or less than about 10 nm or less in some cases. In one embodiment, the channel is a capillary.

A variety of materials and methods, according to certain aspects of the invention, can be used to form devices or components containing microfluidic channels, chambers, etc. For example, various devices or components can be formed from solid materials, in which the channels can be formed via micromachining, film deposition processes such as spin coating and chemical vapor deposition, physical vapor deposition, laser fabrication, photolithographic techniques, etching methods including wet chemical or plasma processes, electrodeposition, and the like. See, for example, *Scientific American*, 248:44-55, 1983 (Angell, *et al*).

In one set of embodiments, various structures or components can be formed of a polymer, for example, an elastomeric polymer such as polydimethylsiloxane ("PDMS"), polytetrafluoroethylene ("PTFE" or Teflon®), or the like. For instance, according to one embodiment, a channel such as a microfluidic channel may be implemented by fabricating the fluidic system separately using PDMS or other soft lithography techniques (details of soft lithography techniques suitable for this embodiment are discussed in the references entitled

"Soft Lithography," by Younan Xia and George M. Whitesides, published in the *Annual Review of Material Science*, 1998, Vol. 28, pages 153-184, and "Soft Lithography in Biology and Biochemistry," by George M. Whitesides, Emanuele Ostuni, Shuichi Takayama, Xingyu Jiang and Donald E. Ingber, published in the *Annual Review of Biomedical Engineering*, 2001, Vol. 3, pages 335-373; each of these references is incorporated herein by reference).

Other examples of potentially suitable polymers include, but are not limited to, polyethylene terephthalate (PET), polyacrylate, polymethacrylate, polycarbonate, polystyrene, polyethylene, polypropylene, polyvinylchloride, cyclic olefin copolymer (COC), polytetrafluoroethylene, a fluorinated polymer, a silicone such as polydimethylsiloxane, polyvinylidene chloride, bis-benzocyclobutene ("BCB"), a polyimide, a fluorinated derivative of a polyimide, or the like. Combinations, copolymers, or blends involving polymers including those described above are also envisioned. The device may also be formed from composite materials, for example, a composite of a polymer and a semiconductor material.

In some embodiments, various microfluidic structures or components of the device are fabricated from polymeric and/or flexible and/or elastomeric materials, and can be conveniently formed of a hardenable fluid, facilitating fabrication via molding (e.g. replica molding, injection molding, cast molding, etc.). The hardenable fluid can be essentially any fluid that can be induced to solidify, or that spontaneously solidifies, into a solid capable of containing and/or transporting fluids contemplated for use in and with the fluidic network. In one embodiment, the hardenable fluid comprises a polymeric liquid or a liquid polymeric precursor (i.e. a "prepolymer"). Suitable polymeric liquids can include, for example, thermoplastic polymers, thermoset polymers, waxes, metals, or mixtures or composites thereof heated above their melting point. As another example, a suitable polymeric liquid may include a solution of one or more polymers in a suitable solvent, which solution forms a solid polymeric material upon removal of the solvent, for example, by evaporation. Such polymeric materials, which can be solidified from, for example, a melt state or by solvent evaporation, are well known to those of ordinary skill in the art. A variety of polymeric materials, many of which are elastomeric, are suitable, and are also suitable for forming molds or mold masters, for embodiments where one or both of the mold masters is composed of an elastomeric material. A non-limiting list of examples of such polymers includes polymers of the general classes of silicone polymers, epoxy polymers, and acrylate polymers. Epoxy polymers are characterized by the presence of a three-membered cyclic ether group commonly referred to as an epoxy group, 1,2-epoxide, or oxirane. For example, diglycidyl

ethers of bisphenol A can be used, in addition to compounds based on aromatic amine, triazine, and cycloaliphatic backbones. Another example includes the well-known Novolac polymers. Non-limiting examples of silicone elastomers suitable for use according to the invention include those formed from precursors including the chlorosilanes such as methylchlorosilanes, ethylchlorosilanes, phenylchlorosilanes, etc.

Silicone polymers are used in certain embodiments, for example, the silicone elastomer polydimethylsiloxane. Non-limiting examples of PDMS polymers include those sold under the trademark Sylgard by Dow Chemical Co., Midland, MI, and particularly Sylgard 182, Sylgard 184, and Sylgard 186. Silicone polymers including PDMS have several beneficial properties simplifying fabrication of various structures of the invention. For instance, such materials are inexpensive, readily available, and can be solidified from a prepolymeric liquid via curing with heat. For example, PDMSs are typically curable by exposure of the prepolymeric liquid to temperatures of about, for example, about 65 °C to about 75 °C for exposure times of, for example, at least about an hour. Also, silicone polymers, such as PDMS, can be elastomeric and thus may be useful for forming very small features with relatively high aspect ratios, necessary in certain embodiments of the invention. Flexible (e.g., elastomeric) molds or masters can be advantageous in this regard.

One advantage of forming structures such as microfluidic structures or channels from silicone polymers, such as PDMS, is the ability of such polymers to be oxidized, for example by exposure to an oxygen-containing plasma such as an air plasma, so that the oxidized structures contain, at their surface, chemical groups capable of cross-linking to other oxidized silicone polymer surfaces or to the oxidized surfaces of a variety of other polymeric and non-polymeric materials. Thus, structures can be fabricated and then oxidized and essentially irreversibly sealed to other silicone polymer surfaces, or to the surfaces of other substrates reactive with the oxidized silicone polymer surfaces, without the need for separate adhesives or other sealing means. In most cases, sealing can be completed simply by contacting an oxidized silicone surface to another surface without the need to apply auxiliary pressure to form the seal. That is, the pre-oxidized silicone surface acts as a contact adhesive against suitable mating surfaces. Specifically, in addition to being irreversibly sealable to itself, oxidized silicone such as oxidized PDMS can also be sealed irreversibly to a range of oxidized materials other than itself including, for example, glass, silicon, silicon oxide, quartz, silicon nitride, polyethylene, polystyrene, glassy carbon, and epoxy polymers, which have been oxidized in a similar fashion to the PDMS surface (for example, via exposure to an oxygen-containing plasma). Oxidation and sealing methods useful in the context of the

present invention, as well as overall molding techniques, are described in the art, for example, in an article entitled "Rapid Prototyping of Microfluidic Systems and Polydimethylsiloxane," *Anal. Chem.*, 70:474-480, 1998 (Duffy *et al.*), incorporated herein by reference.

The following documents are each incorporated herein by reference in their entireties: U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*; and Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*

In addition, incorporated herein by reference in their entireties are U.S. Provisional Patent Application Serial No. 62/031,062, filed July 30, 2014, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/050,636, filed September 15, 2014, entitled "Probe Library Construction," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/142,653, filed April 3, 2015, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; and a PCT application filed on even date herewith, entitled "Probe Library Construction," by Zhuang, *et al.*

The following examples are intended to illustrate certain embodiments of the present invention, but do not exemplify the full scope of the invention.

EXAMPLE 1

The example presents a platform to enable the simultaneous detection of the number and spatial organization of thousands of distinct mRNAs within single cells with high efficiency and low error-rate using a novel form of highly multiplexed fluorescence *in situ* hybridization (FISH). This example accomplishes these measurements by integrating and innovating methods for massively parallel probe synthesis, super-resolution imaging, and self-correcting error-checking codes.

Here, these examples present methods for the simultaneous detection of some or all of the thousands of unique RNAs expressed in a cell. This approach not only promises to revolutionize the throughput of the already effective single-molecule FISH (smFISH) approach, but also allows researchers to benefit from the hypothesis free discovery approach which has made other whole-genome systems approaches to biology so effective. For example, this whole genome approach may allow researchers to discover RNAs whose expression levels and/or subcellular localization patterns change under certain conditions of interest, such as disease states, without knowing, *a priori*, which mRNA will change in

abundance or localization. Simultaneous measurements of hundreds of genes within a single cell also allow for the identification of correlations between genes in expression and localization patterns in some cases.

This can be achieved using methods for highly multiplexed smFISH via the sequential hybridization of orthogonal detection probes and super-resolution imaging, reducing the cost of probe synthesis, and the development of a highly automated system to minimize demands on the user, as discussed herein. This provides an integrated platform to handle the bioinformatics of probe design, the mathematics of error-correcting codes, the complexity of image registration and analysis, and the cumbersome fluid handling through a simple suite of user friendly interfaces. This integration allows easy operation with limited user training and facilitates the rapid collection of data.

This example illustrates: (1) computational design of "codewords" to attach to all RNA targets in the cell that will allow unique identification of each RNA with some degree of experimental error tolerance, (2) translation of these codewords into nucleotide sequences and synthesis of the required single-stranded (ss) oligonucleotide (e.g. ssDNA) probes, (3) sample fixation and hybridization of these probes to the RNA targets *in situ*, (4) read-out of these codewords via successive rounds of hybridization of distinct fluorescent probes imaged with conventional fluorescence microscopy or super-resolution fluorescence microscopy, and (5) automated decoding of measured codewords combined with computational error correction to uniquely and robustly identify individual mRNAs.

In the first step, a "codeword" is assigned to every RNA that is to be labeled. In a typical design these may be strings of N binary letters or positions. Codewords may be chosen from the same wide range of existing error tolerant or error-correcting encoding schemes developed for digital storage and communication, e.g., using Hamming codes or the like. For example, actin-RNA may be assigned the binary codeword 11001010. Each codeword may be unique and separated from the other codewords by a Hamming distance, h, which measures the number of letters or positions that must be incorrectly read for one codeword to be misinterpreted as a different one. A Hamming distance greater than 1 between all codewords allows for some measurement errors to be detected—since simple errors would produce codewords that are not used to encode RNAs. For a Hamming distance larger than 2, it is also possible to correct some errors, as codewords with one error will be closest in Hamming distance to a single, unique codeword. The total number of different RNAs to be detected from the transcriptome and the amount of error correction desired

determines the length of the codewords. Information theory provides several efficient algorithms for assembling error-correcting binary codebooks.

In the second step, this encoding scheme is translated into a set of oligonucleotide (e.g. DNA) probe sequences, which can be called primary probes or encoding probes, each of which not only targets a probe to the RNA of interest but also encodes the unique binary codeword within a set of secondary binding sites (Fig. 1). For example, first designed may be primary binding sequences for each targeted mRNA. These sequences are "target sequences" that are comprised of complementary nucleotide sequences to their target RNAs computationally selected to satisfy a stringent set of hybridization conditions, including uniqueness in the target genome. To improve the efficiency of hybridization to individual mRNAs, multiple primary target sequences are designed for each individual RNA. Then each position within the set of codewords is assigned a unique oligonucleotide (e.g. DNA) sequence, which is called a read sequence. These tags are designed as to have no interaction with endogenous mRNA sequences or each other. For instance, for all the value "1"s in a codeword of an individual mRNA, the corresponding read sequence is attached to the primary targeting sequences against that mRNA. In general, each probe will contain a target sequence and one or more read sequences. If the total length of the necessary read sequences and the primary target sequence exceeds synthesis capabilities, then subsets of the read sequences can be appended to distinct target sequences. For example, consider the potential codeword 11001010 for actin. Probe sequences for this RNA could contain the read sequences corresponding to positions 1, 2, 5, and 7 in the codeword attached to a variety of target sequences specific to actin. After all the sequences have been designed, the resulting complex set of unique custom oligonucleotide (e.g. DNA) sequences is manufactured and amplified using methods as described below.

In the third step, the resulting pool of DNA is hybridized, e.g., to fixed, permeabilized cells. In this process, individual probes may be attached to every RNA in the cell by hybridization of their corresponding target sequences with the RNA while the read sequences remain free to bind the appropriate secondary probes as discussed below.

In the fourth step—the read-out step—fluorescently labeled secondary nucleic acid probes (also called readout probes) are successively hybridized to the read sequences attached to the target sequences that binds to the mRNA targets in the above step. When simultaneously imaging a large number of different RNA species in cells, the density of labeled RNAs may exceed that at which each RNA can be resolved via conventional imaging methods. Thus, this may be performed using a super-resolution imaging method, for example

STORM (stochastic optical reconstruction microscopy), to resolve the labeled molecules. After each round of hybridization and imaging with the secondary probes, the fluorophores are quenched or otherwise inactivated either via chemical or optical techniques such as oxidation, chemically bleaching, photobleaching, stringent washing or enzymatic digestion, etc. The sample is then stained with the next secondary probe, and the cycle continues until all positions of the codewords have been read out. In the simplest incarnation, there will be one hybridization step for each position within the codeword, e.g. 8 hybridization steps for an 8-letter codeword (Fig. 1).

Fig. 1 shows schematic diagrams of this example. Fig. 1A shows that every position of the codewords is assigned a unique oligonucleotide sequence when this position has a value "1." All mRNA codewords are then translated into combinations of read sequences, which are attached to the targeting sequence. Fig. 1B shows various steps of the labeling scheme of this example. In the first step, all mRNAs (I - III) are tagged with multiple oligonucleotide (e.g. ssDNA) probes comprising a primary targeting sequence, which hybridizes to the RNA of interest, and a "tail" (i.e. containing read sequences) carrying the translated codeword, which does not interact with endogenous nucleotide sequences. In the next step, the first secondary probe is added, which can bind all probes whose tails have a read sequence corresponding to the value of "1" in the first position. The dyes on these secondary probes are imaged and bleached, then the next secondary probe is added to bind probes attached to mRNA which have a value of "1" in the second position of their assigned codeword, and so on.

In the final step, the microscopy images from each staining and imaging round are aligned, for example, computationally (e.g. using fiducial beads or other markers tracked during image acquisition), and the clusters of localizations resolved by conventional fluorescence microscopy or super-resolution imaging (e.g. STORM) from the different rounds are identified. These clusters of localizations arise from individual target mRNA molecules, and the hybridization rounds in which a spot was detected in a given cluster correspond to the "1" in the codeword for that mRNA. If there are no missed-detection events or false positive signal in the images, this codeword will perfectly match one of the expected codewords. Fig. 1 describes an example in which the codeword has three letters, i.e. three positions, and the three target mRNAs have codewords 110, 101, and 011 assigned to them. In real experimental examples, the codeword could contain more digits. For example, the mRNA for actin can be assigned the codeword, 11001010. In that case, detected clusters containing overlapping localization signals in the 1st, 2nd, 5th and 7th hybridization steps

(meaning the 1st, 2nd, 5th and 7th secondary probes bound to this site) can be identified as individual actin mRNA molecules, since the pattern of positive bindings matches the codewode of actin (11001010). In addition, if there are missed-detection events or false positive signals in the image data, these aberrations can be corrected by the implemented error-correction scheme. For example, clusters of localizations with a detected codeword that has only one digit discrepancy from 11001010 (such as 11000010 or 11101010) can also be identified as actin mRNA since all other valid codeword in this example differ from the detected pattern in two more positions.

EXAMPLE 2

This example describes another alternative approach that differs in several of the steps described above. This approach begins with the first step, construction of the codewords to the desired mRNA targets, as described above.

In the second step of this approach, nucleic acid probes are designed that bind uniquely to the mRNA targets of interest, as described above. However, instead of appending unique read sequences to these targeting sequences, unique pools or groups of probes are constructed from these target sequences. Each pool comprises all or a subset of the sequences that target all mRNAs which contain the same value at a given position in their codeword. For example, the first pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the first position of their codewords, e.g. 110 and 101 but not 011; the second pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the second position of their codewords, e.g. 110 and 011 but not 101; the third pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the third position of their codewords, e.g. 011 and 101 but not 110 (Fig. 1C). As another example, consider the potential codeword 11001010 for actin. Probes that target this mRNA would be included in pools 1, 2, 5, and 7 but not in pools 3, 4, 6, and 8. The same target to a given mRNA may or may not be included in pools. For example, a probe that targets the same region of actin may be included in pools 1, 2, 5, and 7 or any subset of these pools. After all of the pools have been designed, each complex set of unique, custom oligonucleotide sequences is manufactured and amplified using methods as described below.

In the third step of this approach, the first pool of probes is hybridized, e.g., to fixed, permeabilized cells. In this process, the fluorophores attached to each of the probes in this pool are bound to each of the targets of that pool. The binding of these probes is then determined by fluorescence microscopy. As described above, these images can be collected either via a range of methods including both conventional fluorescence imaging or super-

resolution imaging methods such as STORM. After a round of imaging, the probes from the first pool are inactivated or removed from the sample via the methods described above. This process is then repeated for each successive pool of probes until some or all of the pools have been applied to the sample and imaged such that all positions in the codewords have been read out. In the simplest incarnation, there will be one hybridization and imaging step for each position in the codeword, e.g. 3 rounds of hybridization and imaging for a codeword with 3 positions (Fig 1C) or 8 rounds of hybridization and imaging for a codeword with 8 positions.

The final step of this approach is identical to that described above.

EXAMPLE 3

In this example, 14 genes (PGK1, H3F3B, PKM, ENO1, GPI, EEF2, GNAS, HSPA8, GAPDH, CALM1, RHOA, PPIA, UBA52, and VCP) were encoded using a subset of the (8,4) SECDED code (Figs. 2A-2E). To determine the accuracy of these measurements, the measured abundances of these 14 mRNAs were compared to the abundances measured from bulk RNA-seq of A549 cells (published ENCODE data). Remarkably, it was found that there was excellent agreement between these two measurements, as the transcripts count measured using the sequential hybridization approach correlated with gene expression measured using RNA-seq with a Pearson correlation coefficient r of 0.75 (Fig. 2F). Gene expression from 3 other cells were also measured, and it was found that the gene expression of these 14 genes was highly correlated among the cells with a r of 0.96 (Fig. 2G).

Codebook Design. Each mRNA in the target set was assigned a binary codeword using a Single Error Correction Double Error Detection (SECDED) code. SECDED is an extended Hamming codebook with an additional parity bit. Briefly, Matlab's Communications System toolbox was used to generate SECDED codes of either 8 or 16 letters or positions. In both cases, only those codewords containing four 1s were used. These words were assigned at random to mRNAs in the target set. [0 1 0 1 1 1 0 0] is an example of the 8-letter codewords used (i.e., these codewords each contained four 1s and four 0s.) [0 1 0 1 1 1 0 0 0 0 0 0 0 0 0 0] is an example of the 16-letter codewords used (i.e., each codeword contained four 1s and twelve 0s). Not every codeword was necessarily assigned to an mRNA.

Computational Assembly of ssDNA Primary Probe Sequence. The number of primary nucleic acid probes used for hybridization with mRNA targets ranged from 200 to 2000 unique oligonucelotides, depending on the experiment. For example, to label 14 mRNAs with 28 oligos targeting each gene, 392 unique sequences were used. Large number

of oligos with unique sequences were purchased in a pool from LC Sciences or CustomArray. However, array synthesized oligos were in minute quantity that was insufficient for *in situ* hybridization. The protocol for their amplification are described below.

Each primary probe contained three components: flanking primer sequences to allow enzymatic amplification of probes, targeting sequence for *in situ* hybridization to mRNAs, and secondary tag sequence containing one or more read sequences for sequential readout of codewords.

The following is an example of a primary probe:

GTTGGCGACGAAAGCACTGCGATTGGAACCGTCCCAAGCGTTGCG
CTTAATGGATCATCAATTTTGTCTCACTACGACGGTCAATCGCGCTGCATACTTG
CGTCGGTCGGACAAACGAGG

(SEQ ID NO: 1)

The components are arranged in the following order: forward primer (not underlined), secondary read sequence 1 (underlined), mRNA targeting sequence (not underlined), secondary read sequence 2 (underlined), and reverse primer (not underlined). The secondary read sequences are the reverse complement of the corresponding secondary probes. Since only codewords that contained four '1's were used, the primary probes for each mRNA needed to contain 4 different secondary read sequences in this example. However, in order to reduce the overall length of the primary probes, the pool of targeting sequences for each mRNA target was split at random into two pools. Two secondary read sequences are attached each probe in one of the two pools and the other two secondary read sequences are attached the probes in the other pool. The design criteria for each component are described below.

Primer Design. Specific index primers were generated by a collection of 240,000 published sequences of orthogonal 25-bp long sequences. These sequences were trimmed to 20 bp, selected for a narrow 70 to 80 ^{o}C melting temperature, the absence of consecutive repeats of 3 or more base, and the presence of a GC clamp, i.e. one of the two 3' terminal bases must be G or C. To further improve specificity, these sequences were then screened against the human genome using BLAST+ (Camacho et al 2009), and primers with 14 or more contiguous bases of homology were eliminated. In a subsequent screening via BLAST+, primers that shared 11 or more contiguous bases or more than 5 bases at the 3' end of any other primer or the T7 promoter were also removed.

Secondary Probes Design. 30-bp long secondary probe sequences were created by concatenating fragments of the orthogonal primer set described above. These secondaries were then screened for orthogonality with other secondaries (no more than 11 basepairs of homology) and for potential off-target binding sites in the human genome (no more than 14 basepairs of homology). Secondary sequences used in this example are provided in Table 1.

Table 1

Bit	Secondary sequences	Sequence number
B1	CGCAACGCTTGGGACGGTTCCAATCGGATC	SEQ ID NO: 2
B2	CGAATGCTCTGGCCTCGAACGAACGATAGC	SEQ ID NO: 3
B3	ACAAATCCGACCAGATCGGACGATCATGGG	SEQ ID NO: 4
B4	CAAGTATGCAGCGCGATTGACCGTCTCGTT	SEQ ID NO: 5
B5	TGCGTCGTCTGGCTAGCACGGCACGCAAAT	SEQ ID NO: 6
B6	AAGTCGTACGCCGATGCGCAGCAATTCACT	SEQ ID NO: 7
B7	CGAAACATCGGCCACGGTCCCGTTGAACTT	SEQ ID NO: 8
B8	ACGAATCCACCGTCCAGCGCGTCAAACAGA	SEQ ID NO: 9
B9	CGCGAAATCCCCGTAACGAGCGTCCCTTGC	SEQ ID NO: 10
B10	GCATGAGTTGCCTGGCGTTGCGACGACTAA	SEQ ID NO: 11
B11	CCGTCGTCTCCGGTCCACCGTTGCGCTTAC	SEQ ID NO: 12
B12	GGCCAATGGCCCAGGTCCGTCACGCAATTT	SEQ ID NO: 13
B13	TTGATCGAATCGGAGCGTAGCGGAATCTGC	SEQ ID NO: 14
B14	CGCGCGGATCCGCTTGTCGGGAACGGATAC	SEQ ID NO: 15
B15	GCCTCGATTACGACGGATGTAATTCGGCCG	SEQ ID NO: 16
B16	GCCCGTATTCCCGCTTGCGAGTAGGGCAAT	SEQ ID NO: 17

mRNA Targeting Sequence Design. To determine the relative abundance of all the isoforms of all genes expressed in these cell lines, transcriptome profiling data from the ENCODE project for total RNA from A549 and IMR90 cells was processed using the publicly available software *cufflinks*, along with human genome annotations from gencode v18. Gene models corresponding to the most highly expressed isoform were used to build a sequence library in FASTA format recording the dominant isoform of every gene. Genes of interest were selected from this library. These genes were partitioned into 1kb segments, then

the software *OligoArray2.1* was used to generate primary probe sequences for the human transcriptome with the following constraints: 30-bp or 40-bp length, depending on the experiment; probe-target melting temperatures greater than 70 °C (variable parameter); no cross hybridization targets with melting temperatures greater than 72 °C (variable parameter); no predicted internal secondary structures with melting temperatures greater than 76 °C (variable parameter); and no single-nucleotide contiguous repeats of 6 or more bases. After OligoArray probe selection, all potential probes that mapped to a different gene were rejected while all potential probes with multiple alignments to the same gene were retained. A BLAST database was assembled from the FASTA library of all expressed genes to screen for probes' uniqueness. For each gene, 14 to 28 targeting sequences produced during the OligoArray processing were selected.

Probe Synthesis – index PCR. The template for specific probe sets were selected from the complex oligopool via limited-cycle PCR. Briefly, 0.5 to1 ng of the complex oligopool was combined with 0.5 micromolar of each primer. The forward primer matched the priming sequence for the desired subset while the reverse primer was a 5' concatenation of this sequence with a T7 promoter. To avoid the generation of G-quadruplets, which can be difficult to synthesize, the terminal Gs required in the T7 promoter were generated from Gs located at the 5' of the priming region where appropriate. All primers were synthesized by IDT. A 50 microliter reaction volume was amplified either using the KAPA real-time library amplification kit (KAPA Biosystems; KK2701) or via a homemade qPCR mix which included 0.8X EvaGreen (Biotum; 31000-T) and the hot-start Phusion polymerase (New England Biolabs; M0535S). Amplification was followed in real time using Agilent's MX300P or Biorad's CFX Connect. Individual samples were removed immediately before the plateau in amplification to minimize distortion of template abundance due to over-amplification. Individual templates were purified with columns according to the manufacturer's instructions (Zymo DNA Clean and Concentrator; D4003) and eluted in RNase-free deionized water.

Amplification via *in-vitro* transcription. The template was then amplified via *in vitro* transcription. Briefly, 0.5 to 1 microgram of template DNA was amplified into 100-200 microgram of RNA in a single 20-30 microliter reaction with a high yield RNA polymerase (New England Biolabs; E2040S). Reactions were supplemented with 1X RNase inhibitor (Promega RNasin; N2611). Amplification was typically run for 4 to 16 hours at 37 °C to maximize the yield. The RNA was not purified after the reaction and was either stored at -80 °C or immediately converted into DNA as described below.

Reverse Transcription. 1 to 2 nmol of fluorescently-labeled ssDNA probe was created from the above *in vitro* transcription reactions using the reverse transcriptase Maxima H- (Thermo Scientific; EP0751). This enzyme was used because of its higher processivity and temperature resistance, which allowed the conversion of large quantities of RNA into DNA within small volumes at temperatures that disfavor secondary structure formation. The unpurified RNA created above was supplemented with 1.6 mM of each dNTP, 1-2 nmol of fluorescently labeled forward primer, 300 units of Maxima H-, 60 units of RNasin, and a final 1X concentration of the Maxima RT buffer. The final 75 microliter volume was incubated at 50 °C for 60 minutes.

Strand Selection and Purification. The template RNA in the reaction above was then removed from the DNA via alkaline hydrolysis. 75 microliters of 0.25 M EDTA and 0.5 N NaOH were added to each reverse transcription reaction, and the sample was incubated at 95 °C for 10 minutes. The reaction was immediately neutralized by purifying the ssDNA probe with a modified version of the Zymo Oligo Clean and Concentrator protocol. Specifically, the 5-microgram capacity column was replaced with a 25-microgram or 100-microgram capacity DNA column as appropriate. The remainder of the protocol was run according to the manufacturer's instructions. Probe was eluted in 100 microliter RNase-free deionized water and evaporated in a vacuum concentrator. The final pellet was resuspended in 10 microliters of RNase-free water and stored at -20 °C. Denaturing poly-acrylimide gel electrophoresis and absorption spectroscopy revealed that this protocol typically produced 90-100% incorporation of the fluorescent primer into full length probe and 75-90% recovery of the total fluorescent probe. Thus, without exceeding a 150-microliter reaction volume, this protocol can be used to create ~2 nmol of fluorescent probe.

Cell culture and fixation. A549 and IMR90 cells (American Type Culture Collection) were cultured with Dulbecco's Modified Eagle Medium and Eagle's Minimum Essential Medium respectively. Cells were incubated at 37 °C with 5% CO_2 for 36-48 hours. Cells were fixed in 3% paraformaldehyde (Electron Microscopy Sciences) in PBS for 15 minutes, washed with PBS, and permeabilized in 70% ethanol overnight at 4 °C.

Fluorescence In Situ Hybridization (FISH) – primary (encoding) probes. Cells were hydrated in wash buffers (2xSSC, 50% formamide) for 10 minutes, labelled with primary oligos (0.5 nM per sequence) in hybridization buffers (2xSSC, 50% formamide, 1 mg/mL yeast tRNA, and 10% dextran sulfate) overnight at 37 °C, washed with wash buffers at 47 °C for 10 minutes twice, and washed with 2xSSC twice. Fluorescent fiducial beads (Molecular Probes, F-8809) were added at a 1:10,000 dilution in 2xSSC before imaging.

Secondary probes. Secondary (readout) probes (10 nM) were hybridized in secondary hybridization buffers (2xSSC, 20% formamide, and 10% dextran sulfate) to their primary targets for 30 minutes at 37 °C. Cells remained on the microscope stage during the hybridization. An objective heater was used to maintain the temperature at 37 °C. Cells were washed with secondary wash buffers (2xSSC, 20% formamide).

Fluidics and STORM Imaging. Multiple rounds of sequential labeling, washing, imaging, and bleaching were performed on an automated platform consisted of a fluidics setup and a STORM (stochastic optical reconstruction microscopy) microscope. The fluidics setup included a flow chamber (Bioptech FCS2), a peristaltic pump (Rainin Dynamax RP-1), and three computer-controlled 8-way valves (Hamilton MVP and Hamilton HVXM 8-5). This system allowed the automated integration of STORM movie collection and secondary hybridization.

The imaging buffer included, 50 mM Tris (pH 8) 10% (w/v) glucose, 1% βME (2-mercaptoethanol) or 25 mM MEA, with or without 2 mM 1,5-cyclooctadiene, and an oxygen scavenging system (0.5 mg/ml glucose oxidase (Sigma-Aldrich) and 40 microgram/ml catalase (Sigma-Aldrich)). A layer of mineral oil was used to seal the imaging buffer, preventing its acidification over the course of multiple hybridizations.

The STORM setup included an Olympus IX-71 inverted microscope configured for oblique incidence excitation. The samples were continuously illuminated with a 642-nm diode-pumped solid-state laser (VFL-P500-642; MPB communications). A 405-nm solid-state laser (Cube 405-100C; Coherent) was used for activation of dyes. Fluorescence was collected using an Olympus (UPlanSApo 100x, 1.4 NA) objective lens and passed through a custom dichroic, as well as a quad-view beam splitter. All movies were recorded using an EMCCD camera (Andor iAxon 897), imaging at 60Hz. The 512x256 field of view of the camera was split into separate 256x256 pixel movies prior to saving. The left half of this field of view contained the STORM data and the right half contained images of the fluorescent feducial beads. These latter movies were downsampled to 1 Hz prior to saving. During data acquisition, a home-built focus lock was used to maintain a constant focal plane. STORM movies included 20,000 to 30,000 frames in STORM buffer while the bleach movies included 10,000 frames in wash buffer.

Image Analysis – analysis of single-molecule localizations. Movies of single molecule localizations and fluorescent feducial beads were processed separately using a previously published single-emitter localization software.

Image Registration. The starting position of the beads from each round of hybridization were used to align movies from each round. The 2D autocorrelation between bead images of consecutive hybridizations followed by nearest-neighbor matching was used to match beads between images. The pair of beads with the most similar displacement vector were used to compute a rigid translation-rotation warp to align the beads. This alignment method is robust to samples in which multiple feducials are displaced or come detached and reattach during imaging.

Drift Correction. Drift during image acquisition was corrected using the trajectory of the feducial beads (recorded at 1 Hz). Bead positions were linked in each frame. The trajectory of the two beads that moved in the most correlated fashion was taken as the drift trajectory.

mRNA Cluster Calling. Localizations were first screened to be above a threshold number of photons (generally 2000) and required to be within 32 nm of 5 other localizations (parameters may be tuned). The remaining molecule localizations were binned in a 2D histogram of 10x10 nm bins (bin size is a variable parameter). All connected bins were taken to be part of a cluster (diagonal contacts are classified as connected). Clusters were required to have more than 80 total localizations across all hybridizations (variable parameter) to be called an mRNA cluster. The weighted centroids of these clusters from the 2D histogram were recorded as the mRNA positions.

A given cluster is recorded to be represented in an individual hybridization round if more than 9 localizations (variable parameter) are found within a 48 nm radius (variable parameter) of the centroid for that mRNA in each hybridization round.

Cluster Decoding. For each mRNA cluster, a codeword is readout, including "0"s for all the hybridization rounds in which less than the threshold number of localizations are found near the centroid and "1" s for the rounds where above threshold number of localizations are counted. The SECDED codebook decoded these as either perfect matches to target mRNA codewords, correctable errors which can be unambiguously mapped back to target mRNA, or uncorrectable errors, which differed by two or more letters from the words in the codebook.

Fig. 2A shows a STORM image of a cell. Fig. 2B shows a zoom in of the boxed region in Fig. 2A. Each dot indicates a localization. Localizations from different rounds of imaging are shown differently. Fig. 2C shows a representative cluster of localizations from the boxed region in Fig. 2B. The cluster shows localization signals from 4 different hybridizations. This cluster is a putative mRNA encoded with codeword [0 1 0 1 1 1 0 0].

Fig. 2D shows a reconstructed cell image of 14 genes after decoding and error correction. Each gene is shown differently. Fig. 2E shows measured gene expression for the 14 genes from the cell. Fig. 2F shows a comparison of transcript count with ensemble RNA-sequencing data. Fig. 2G shows correlation of transcript expression level between two cells detected using the described approach.

EXAMPLE 4

The following examples are generally directed to multiplexed single-molecule imaging with error-robust encoding allowing for simultaneous measurements of thousands of RNA species in single cells. In general, knowledge of the expression profile and spatial landscape of RNAs in individual cells is essential for understanding the rich repertoire of cellular behaviors. The following examples reports various techniques directed to single-molecule imaging approaches that allow the copy numbers and spatial localizations of thousands of RNA species to be determined in single cells. Some of these techniques are called Multiplexed Error-Robust Fluorescence in Situ Hybridization or "MERFISH."

Using error-robust encoding schemes to combat single-molecule labeling and detection errors, these examples demonstrated the imaging of hundreds to thousands of unique RNA species in hundreds of individual cells. Correlation analysis of the ~10^4 to ~10^6 pairs of genes allowed constraints on gene regulatory networks, prediction of novel functions for many unannotated genes, and identification of distinct spatial distribution patterns of RNAs that correlate with properties of the encoded proteins.

System-wide analyses of the abundance and spatial organization of RNAs in single cells promise to transform understanding in many areas of cell and developmental biology, such as the mechanisms of gene regulation, the heterogeneous behavior of cells, and the development and maintenance of cell fate. Single-molecule fluorescence *in situ* hybridization (smFISH) has emerged as a powerful tool for studying the copy number and spatial organization of RNAs in single cells either in isolation or in their native tissue context. Taking advantage of its ability to map the spatial distributions of specific RNAs with high resolution, smFISH has revealed the importance of subcellular RNA localization in diverse processes such as cell migration, development, and polarization. In parallel, the ability of smFISH to precisely measure the copy numbers of specific RNAs without amplification bias has allowed quantitative measurement of the natural fluctuations in gene expression, which has in turn elucidated the regulatory mechanisms that shape such fluctuations and their role in a variety of biological processes.

However, application of the smFISH approach to many systems-level questions remains limited by the number of RNA species that can be simultaneously measured in single cells. State-of-the-art efforts using combinatorial labeling by either color-based barcodes or sequential hybridization have enabled simultaneous measurements of 10-30 different RNA species in individual cells, yet many interesting biological questions would benefit from the measurement of hundreds to thousands of RNAs within a single cell, which are not achievable using such techniques. For example, analysis of how the expression profile of such a large number of RNAs vary from cell to cell and how these variations correlate among different genes could be used to systematically identify co-regulated genes and map regulatory networks; knowledge of the subcellular organizations of numerous RNAs and their correlations could help elucidate molecular mechanisms underlying the establishment and maintenance of many local cellular structures; and RNA profiling of individual cells in native tissues could allow *in situ* identification of cell type.

The following examples generally discuss certain techniques called MERFISH, which are highly multiplexed smFISH imaging methods that substantially increase the number of RNA species that can be simultaneously imaged in single cells by using combinatorial labeling and sequential imaging with error-robust encoding schemes. These examples demonstrate this multiplexed imaging approach by simultaneously measuring 140 RNA species using an encoding scheme that can both detect and correct errors and 1001 RNA species using an encoding scheme that can detect but not correct errors. It should be understood that these numbers are by way of exemplification only, not limitation. Correlation analyses of the copy number variations and spatial distributions of these genes allowed us to identify groups of genes that are co-regulated and groups of genes that share similar spatial distribution patterns inside the cell.

Combinatorial labeling with error-robust encoding schemes. Combinatorial labeling that identifies each RNA species by multiple (N) distinct signals offers a route to rapidly increase the number of RNA species that can be probed simultaneously in individual cells (Fig. 5A). However, this approach to scaling up the throughput of smFISH to the systems scale faces a significant challenge because not only does the number of addressable RNA species increases exponentially with N, but the detection error rates also grow exponentially with N (Figs. 5B-5D). Imagine a conceptually simple scheme to implement combinatorial labeling, where each RNA species is encoded with a N-bit binary word and the sample is probed with N corresponding rounds of hybridization, each round targeting only the subset of RNAs that should read '1' in the corresponding bit (Fig. 11). N rounds of hybridization

would allow 2^N-1 RNA species to be probed. With just 16 hybridizations, over 64,000 RNA species, which should cover the entire human transcriptome including both messenger RNAs (mRNAs) and non-coding RNAs, could be identified (Fig. 5B; upper symbols). However, as *N* increases, the fraction of RNAs properly detected (the calling rate) would rapidly decrease and, more troublingly, the fraction of RNAs that are identified as incorrect species (the misidentification rate) would rapidly increase (Fig. 5C, lower symbols; Fig. 5D, upper symbols). With realistic error rates per hybridization (measured below), the majority of RNA molecules would be misidentified after 16 rounds of hybridizations!

To address this challenge, error-robust encoding schemes were designed, in which only a subset of the 2^N-1 words separated by a certain Hamming distance were used to encode RNAs. In a codebook where the minimum Hamming distance is 4 (HD4 code), at least four bits must be read incorrectly to change one code word into another (Fig. 12A). As a result, every single-bit error produces a word that is uniquely close to a single code word, allowing such errors to be detected and corrected (Fig. 12B). Double-bit errors produce words with an equal Hamming distance of 2 from multiple code words and, thus, can be detected but not corrected (Fig. 12C). Such a code should substantially increase the calling rate and reduce the misidentification rate (Fig. 5C and 5D, middle symbols). To further account for the fact that it is more likely to miss a hybridization event (an 1-->0 error) than to misidentify a background spot as an RNA (an 0-->1 error) in smFISH measurements, a modified HD4 (MHD4) code was designed, in which the number of '1' bits were kept both constant and relatively low, only four per word, to reduce error and avoid biased detection. This MHD4 code should further increase the calling rate and reduce the misidentification rate (Fig. 5C, upper symbols; Fig. 5D, lower symbols).

In addition to the error considerations, several practical challenges have also made it difficult to probe a large number of RNA species, such as the high cost of the massive number of fluorescently labeled FISH probes needed and the long time required to complete many rounds of hybridization. To overcome these challenges, in this example, a two-step labeling scheme was designed to encode and readout cellular RNAs (Fig. 5E). First, cellular RNAs were labeled with a set of encoding probes (also called primary probes), each probe comprising a RNA targeting sequence and two flanking readout sequences. Four of the *N* unique readout sequences were assigned to each RNA species based on the MHD4 code word of the RNA. Second, these *N* readout sequences were identified with complementary FISH probes, the readout probes (also called secondary probes) via *N* rounds of hybridization and imaging, each round using a unique readout probe. To increase the signal to background

ratio, every cellular RNA was labeled with ~192 encoding probes. Because each encoding probe contained two of the four readout sequences associated with that RNA (Fig. 5E), a maximum of ~96 readout probes can bind to each cellular RNA per hybridization round. To generate the massive number of encoding probes required, they were amplified from array-derived oligonucleotide pools containing tens of thousands of custom sequences using an enzymatic amplification process comprising *in vitro* transcription followed by reverse transcription (Fig. 13, see below regarding probe synthesis). This two-step labeling approach significantly diminished the total hybridization time for an experiment: it was found that efficient hybridization to the readout sequences took only 15 minutes whereas efficient direct hybridization to cellular RNA required more than 10 hours.

Fig. 5 describes MERFISH, a highly multiplexed smFISH approach using combinatorial labeling and error-robust encoding. Fig. 5A shows a schematic depiction of the identification of multiple RNA species in N rounds of imaging. Each RNA species is encoded with a N-bit binary word and during each round of imaging, only the subset of RNAs that should read '1' in the corresponding bit emit signal. Figs. 5B-5D show the number of addressable RNA species (Fig. 5B), the rate at which these RNAs are properly identified (calling rate) (Fig. 5C), and the rate at which RNAs are incorrectly identified as a different RNA species (misidentification rate) (Fig. 5D) plotted as a function of the number of bits (N) in the binary words encoding RNA. In Figs. 5B and 5D, the upper dots are a simple binary code that includes all 2^N-1 possible binary words; the middle dots are the HD4 code where the Hamming distance separating words is 4; and the lower dots are the modified HD4 (MHD4) code where the number of '1' bits are kept at four. These are reversed in Fig. 5C.

The calling and misidentification rates are calculated with per bit error rates of 10% for the 1-->0 error and 4% for the 0-->1 error. Fig. 5E is as schematic diagram of the implementation of a MHD4 code for RNA identification. Each RNA species is first labeled with ~192 encoding probes that convert the RNA into a unique combination of readout sequences (Encoding hyb). These encoding probes each contain a central RNA targeting region flanked by two readout sequences, drawn from a pool of N different sequences, each associated with a specific hybridization round. Encoding probes for a specific RNA species contain a unique combination of four of the N readout sequences, which correspond to the four hybridization rounds where this RNA should read '1'. N subsequent rounds of hybridization with the fluorescent readout probes were used to probe the readout sequences (hyb 1, hyb 2, ..., hyb N). The bound probes were inactivated by photobleaching between

successive rounds of hybridization. For clarity, only one possible pairing of the readout sequences is depicted here for the encoding probes; however, all possible pairs of the four readout sequences are used at the same frequency and distributed randomly along each cellular RNA in the actual experiments.

Fig. 11 shows a schematic description of a combinatorial labeling approach based on a simple binary code. In a conceptually simple labeling approach, 2^N-*1* different RNA species can be uniquely encoded with all *N*-bit binary words (excluding the word with all '0's). In each hybridization round, FISH probes that are targeted to all RNA species that have a '1' in the corresponding bit are included. To increase the ability to discriminate RNA spots from background, each RNA is addressed with multiple FISH probes per hybridization round. Signal from the bound probes is extinguished before the next round of hybridization. This process continues for all *N* hybridization rounds (hyb 1, hyb 2, ...), and all 2^N-*1* RNA species can be identified by the unique on-off pattern of fluorescence signals in each hybridization round.

Fig. 12 shows schematic descriptions of Hamming distance and its use in the identification and correction of errors. Fig. 12A is a schematic representation of a Hamming distance of 4. Figs. 12B and 12C are schematics showing the ability of an encoding scheme with Hamming distance 4 to correct single-bit errors (Fig. 12B) or detect but not correct double-bit errors (Fig. 12C). Arrows highlight bits at which the indicated words differ. Two code words are separated by a Hamming distance of 4 if one of the words has to flip four bits from '1' to '0' or '0' to '1' to convert into the other word. Single-bit error correction is possible because if a measured word differs from a legitimate code word by only one bit, it is most likely an error that arises from misreading this code word, since the code words of all the other RNA species will differ from the measured word by at least three bits. In this case, the measured word can be corrected to a code word that differs by only one bit. If a measured word differs from a legitimate code word by two bits, this measured word can still be identified as an error, but correction is no longer possible since more than one legitimate code word differs from this measured word by two bits.

Fig. 13 shows production of the library of encoding probes. An array-synthesized complex oligopool, containing ~100k sequences, is used as a template for the enzymatic amplification of the encoding probes for different experiments. Each template sequence in the oligopool contains a central target region that can bind to a cellular RNA, two flanking readout sequences, and two flanking index primers. In the first step, the required template molecules for a specific experiment are selected and amplified with an indexed PCR reaction.

To allow amplification via *in vitro* transcription, a T7 promotor is added to the PCR products during this step. In the second step, RNA is amplified from these template molecules via *in vitro* transcription. In the third step, this RNA is reverse transcribed back into DNA. In the final step, the template RNA is removed via alkaline hydrolysis, leaving only the desired ssDNA probes. This protocol produces ~2 nmol of complex pools of encoding probes containing ~20,000 different sequences for the 140-gene experiments or ~100,000 different sequences for the 1001-gene experiments.

EXAMPLE 5

This example illustrates the measurement of 140 genes with MERFISH using a 16-bit MHD4 Code. To test the feasibility of this error-robust, multiplexed imaging approach, this example uses a 140-gene measurement on human fibroblast cells (IMR90) using a 16-bit MHD4 code to encode 130 RNA species while leaving 10 code words as misidentification controls (Figs. 20A to 20H). After each round of hybridization with the fluorescent readout probes, cells were imaged by conventional wide-field imaging with an oblique-incidence illumination geometry. Fluorescent spots corresponding to individual RNAs were clearly detected and were then efficiently extinguished via a brief photobleaching step (Fig. 6A). The sample was stable throughout the 16 rounds of iterative labeling and imaging. The change in the number of fluorescent spots from round to round matched the predicted change based on the relative abundances of RNA species targeted in each round derived from bulk sequencing, and a systematic decreasing trend with increasing number of hybridization rounds was not observed (Fig. 14A). The average brightness of the spots varied from round to round with a standard deviation of 40%, likely due to different binding efficiencies of the readout probes to the different readout sequences on the encoding probes (Fig. 14B). Only a small, systematic decreasing trend in the spot brightness with increasing hybridization rounds was observed, which was on average 4% per round (Fig. 14B).

Next, binary words were constructed from the observed fluorescent spots based on their on-off patterns across the 16 hybridization rounds (Fig. 6B-6D). If the word exactly matched one of the 140 MHD4 code words (exact matches) or differed by only one bit (error-correctable matches), it was assigned to the corresponding RNA species (Fig. 6D). Within the single cell depicted in Figs. 6A and 6B, more than 1500 RNA molecules corresponding to 87% of the 130 encoded RNA species were detected after error correction (Fig. 6E). Similar observations were made in ~400 cells from 7 independent experiments. On average, ~4 times as many RNA molecules and ~2 times as many RNA species were detected per cell after error correction as compared with the values obtained before error correction (Fig. 15).

Two types of errors can occur in the copy number measurement of each RNA species: 1) Some molecules of this RNA species are not detected, leading to a drop in calling rate, and 2) some molecules from other RNA species are misidentified as this RNA species. To assess the extent of misidentification, the 10 misidentification control words were utilized, i.e., code words that were not associated with any cellular RNA. Although matches to these control words were observed, they occurred far less frequently than the real RNA-encoding words: 95% of the 130 RNA-encoding words were counted more frequently than the median count for these control words. Moreover, it was typically found that the ratio of the number of exact matches to the number of matches with one-bit errors for a real RNA-encoding word was substantially higher than the same ratios observed for the misidentification controls, as expected (Figs. 16A and 16B). Using this ratio as a measure of the confidence in RNA identification, it was found that 91% of the 130 RNA species had a confidence ratio greater than the maximum confidence ratio observed for the misidentification controls (Fig. 6F), demonstrating a high accuracy of RNA identification. Subsequent analyses were conducted only on these 91% of genes.

To estimate the calling rate, the error-correction ability of the MHD4 code was utilized to determine the 1-->0 error rates (10% on average) and 0-->1 error rates (4% on average) for each hybridization round (Figs. 16C and 16D). Using these error rates, an ~80% calling rate for individual RNA species after error correction was estimated, i.e. ~80% of the fluorescent spots corresponding to a RNA species were decoded correctly (Fig. 16E). It is noted that although the remaining 20% of spots contributed to a loss in detection efficiency, most of them did not cause species misidentification because they were decoded as double-bit error words and discarded.

To test for potential technical bias in these measurements, the same 130 RNAs species were probed with a different MHD4 codebook by shuffling the code words among different RNA species (Figs. 20A to 20H) and changing the encoding probe sequences. Measurements with this alternative code gave similar misidentification and calling rates (Fig. 17). The copy numbers of individual RNA species per cell measured with these two codebooks showed excellent agreement with a Pearson correlation coefficient of 0.94 (Fig. 6G), indicating that the choice of encoding scheme did not bias the measured counts.

In order to validate the copy numbers derived from the MERFISH experiments, conventional smFISH measurements were performed on 15 of the 130 genes, selected from the full measured abundance range of three orders of magnitude. For each of these genes, both the average copy number and the copy number distribution across many cells agreed

quantitatively between the MERFISH and conventional smFISH measurements (Fig. 18A and 18B). The ratio of the copy numbers determined by these two approaches was 0.82 +/- 0.06 (mean +/- SEM across the 15 measured RNA species, Fig. 18B), which agreed with the estimated 80% calling rate for the multiplexed imaging approach. The quantitative match between this ratio and the estimated calling rate over the full measured abundance range additionally supports the assessment that the misidentification error was low. Given that the agreement between the MERFISH and conventional smFISH results extended to the genes at the lowest measured abundance (<1 copy per cell, Fig. 18B), it was estimated that the measurement sensitivity was at least 1 copy per cell.

As a final validation, the abundance of each RNA species averaged over hundreds of cells was compared to those obtained from a bulk RNA sequencing measurement that were performed on the same cell line. The imaging results correlated remarkably well with bulk sequencing results with a Pearson correlation coefficient of 0.89 (Fig. 6H).

Fig. 6 shows simultaneous measurement of 140 RNA species in single cells using MERFISH with a 16-bit MHD4 code. Fig. 6A shows images of RNA molecules in an IMR90 cell after each hybridization round (hyb 1 – hyb 16). The image after photobleaching (bleach 1) demonstrated efficient removal of fluorescent signals between hybridizations. Fig. 6B shows the localizations of all detected single molecules in this cell colored based on their measured binary words. Inset: the composite fluorescent image of the 16 hybridization rounds for the boxed sub-region with numbered circles indicating potential RNA molecules. A circle indicates an unidentifiable molecule, the binary word of which does not match any of the 16-bit MHD4 code words even after error correction. Fig. 6C shows fluorescent images from each round of hybridization for the boxed sub-region in Fig. 6B with circles indicating potential RNA molecules. Fig. 6D shows corresponding words for the spots identified in Fig. 6C. Crosses represent the corrected bits. Fig. 6E shows the RNA copy number for each gene observed without (lower) or with (higher) error correction in this cell. Fig. 6F shows the confidence ratio measured for the 130 RNA species (left) and the 10 misidentification control words (right) normalized to the maximum value observed from the misidentification controls (dashed line). Fig. 6G is a scatter plot of the average copy number of each RNA species per cell measured with two shuffled codebooks of the MHD4 code. The Pearson correlation coefficient is 0.94 with a p-value of 1×10^{-53}. The dashed line corresponds to the $y = x$ line. Fig. 6H is a scatter plot of the average copy number of each RNA species per cell versus the abundance determined by bulk sequencing in fragments per kilobase per million reads

(FPKM). The Pearson correlation coefficient between the logarithmic abundances of the two measurements was 0.89 with a p-value of $3x10^{-39}$.

Fig. 14 shows the number and average brightness of the fluorescent spots detected in the 16 rounds of hybridizmation before and after photobleaching. Fig. 14A shows the number of fluorescent spots observed per cell before (higher) and after (lower) photobleaching as a function of hybridization round averaged across all measurements with the first 16-bit MHD4 code. Photobleaching reduces the number of fluorescent spots by two or more orders of magnitude. Hybridization rounds without lower bars represent rounds in which no molecules were observed after bleaching. Also depicted is the expected change in the number of fluorescent spots from round to round (circles) predicted based on the relative abundances of the RNA species targeted in each hybridization round derived from bulk RNA sequencing. The average discrepancy between the observed and predicted number of spots for each hybridization is only 15% of the mean number of spots. This discrepancy does not systematically increase with the number of hybridization rounds. Fig. 14B shows the average brightness of the identified fluorescent spots in each hybridization round averaged across all measurements with the first 16-bit MHD4 code both before (upper) and after (lower) photobleaching. Brightness varies by 40% (standard deviation) across different hybridization rounds. The variation pattern is reproducible between experiments with the same code, likely due to differences in the binding efficiency of the readout probes to the different readout sequences. There is a small systematic trend of decrease in the brightness with increasing hybridization rounds, which is on average 4% per round. Photobleaching extinguishes fluorescence to a level similar to that of the autofluorescence of the cell.

Fig. 15 shows error correction substantially increases the numbers of RNA molecules and RNA species detected in individual cells. Fig. 15A shows a histogram of the ratio of the total number of molecules detected per cell with error correction to the number measured without error correction. Fig. 15B is a histogram of the total number of RNA species detected in each cell with error correction to that without error correction. Both ratios are determined for ~200 cells and the histograms are constructed from these ratios.

Fig. 16 shows characterization of the misidentification and calling rates of RNA species for the 140-gene experiments using a specific 16-bit MHD4 code. Fig. 16A shows the number of measured words exactly matching the code word corresponding to FLNA, represented by the bar in the center of the circle, and the number of measured words with one-bit error compared to the code word of FLNA, represented by the 16 bars on the circle. Fig. 16B is the same as Fig. 16A, but for a code word that was not assigned to any RNA, i.e.,

a misidentification control word. The solid lines connect the exact match to one-bit error words that are generated by 1-->0 errors. Based on the observation that the ratio of the number of exact matches to the number of error-correctable matches for a real RNA-encoding word was typically substantially higher than the same ratios observed for the misidentification controls, this ratio was defined as a confidence ratio for RNA identification. The confidence ratio measured for all 130 RNA species (center bar) and 10 misidentification control words not assigned to any RNA (outer bars) using this 16-bit MHD4 code is show in Fig. 6F. Fig. 16C and 16D show the average error rates for the 1-->0 error (Fig. 16C) and 0-->1 error (Fig. 16D) for each hybridization round. Fig. 16E shows the calling rate for each RNA species estimated from the 1-->0 and 0-->1 error rates. Genes are sorted from left to right based on the measured abundance, which spans three orders of magnitude. The calling rates are largely independent of the abundance of the gene.

Fig. 17 shows characterization of the misidentification and calling rates for a second 16-bit MHD4 code. In this second encoding scheme, the 140 code words were shuffled among different RNA species and changed the encoding probe sequences. Fig. 17A shows the normalized confidence ratio measured for the 130 RNA species (left) and the 10 misidentification control words not assigned to any RNA (right). The normalized confidence ratio is determined the same way as in Fig. 6F. Figs. 17B and 17C show the average error rates determined for the 1-->0 error (Fig. 17B) and 0-->1 error (Fig. 17C) for each hybridization round. Fig. 17D shows the calling rate determined for each RNA species estimated from the 1-->0 and 0-->1 error rates. Genes are sorted from left to right based on the measured abundance.

Fig. 18 shows a comparison of the MERFISH measurements with conventional smFISH results for a subset of genes. Fig. 18A shows the distributions of RNA copy numbers in single cells for three example genes KIAA1199, DYNC1H1, and LMTK2 in the high, medium, and low abundance ranges, respectively. Lighter bars: distributions constructed from ~400 cells in the 140-gene measurements using the MHD4 codes. Darker bars: distributions constructed from ~100 cells in the conventional smFISH measurements. Fig. 18B shows a comparison of the average RNA copy numbers per cell measured in the 140-gene experiments using the MHD4 codes to those determined by conventional smFISH for 15 genes. The average ratio of the copy number measured using the MHD4 measurements to that measured using conventional smFISH was 0.82 +/- 0.06 (mean +/- SEM across 15 genes). The dashed line corresponds to the $y = x$ line.

Figs. 20A to 20H show[[s]] two different codebooks for the 140-gene experiments. The specific code words of the 16-bit MHD4 code assigned to each RNA species studied in the two shuffles of the 140-gene experiment. The "Genes" columns contain the name of the gene. The "Codewords" columns contain the specific binary word assigned to each gene.

EXAMPLE 6

This example is generally directed to high-throughput analysis of cell-to-cell variation in gene expression. The MERFISH approach allows parallelization of measurements of many individual RNA species and co-variation analysis between different RNA species. In this example, the parallelization aspect was first illustrated by examining the cell-to-cell variation in the expression level of each of the measured genes (Fig. 7A). To quantify the measured variation, the Fano factors, defined as the ratio of the variance to the mean RNA copy number, were computed for all measured RNA species. The Fano factors substantially deviated from 1, the value expected for a simple Poisson process, for many genes and exhibited an increasing trend with the mean RNA abundance (Fig. 7B). This trend of increasing Fano factors with mean RNA abundance can be explained by changes in the transcription rate and/or promoter off-switching rates but not by changes in the promoter on-switching rate.

Moreover, several RNA species were identified with substantially larger Fano factors than this average trend. For example, it was found that SLC5A3, CENPF, MKI67, TNC and KIAA1199 displayed Fano factor values substantially higher than those of the other genes expressed at similar abundance levels. The high variability of some of these genes can be explained by their association with the cell cycle. For example, two of these particularly 'noisy' genes MKI67 and CENPF were both annotated as cell-cycle related genes, and based on their bimodal expression (Fig. 7C), it is proposed that their transcription is strongly regulated by the cell cycle. Other high-variability genes did not show the same bimodal expression patterns and are not known to be associated with the cell cycle.

Analysis of co-variations in the expression levels of different genes can reveal which genes are co-regulated and elucidate gene regulatory pathways. At the population level, such analysis often requires the application of external stimuli to drive gene expression variation; hence, correlated expression changes can be observed among genes that share common regulatory elements influenced by the stimuli. At the single-cell level, one can take advantage of the natural stochastic fluctuations in gene expression for such analysis and can thus study multiple regulatory networks without having to stimulate each of them

individually. Such co-variation analysis can constrain regulatory networks, suggest new regulatory pathways, and predict function for unannotated genes based on associations with co-varying genes.

This approach was applied to the 140-gene measurements and the ~10,000 pairwise correlation coefficients describing how the expression levels of each pair of genes co-varied from cell to cell were examined. Many of the highly variable genes showed tightly correlated or anti-correlated variations (Fig. 7C). To better understand the correlations for all gene pairs, a hierarchical clustering approach was adopted to organize these genes based on their correlation coefficients (Fig. 7D). From the cluster tree structure, seven groups of genes with substantially correlated expression patterns were identified (Fig. 7D). Within each of the seven groups, every gene showed significantly stronger average correlation with other members of the group than with genes outside the group. To further validate and understand these groups, gene ontology (GO) terms enriched in each of these seven groups were identified. Notably, the enriched GO terms within each group shared similar functions and were largely unique to each group (Fig. 7E), validating the notion that the observed co-variation in expression reflects some commonalities in the regulation of these genes.

This example describes two of these groups as illustrative examples. The predominant GO terms associated with Group 1 were terms associated with the extracellular matrix (ECM) (Figs. 7D and 7E). Notable members of this group included ECM components, such as FBN1, FBN2, COL5A, COL7A and TNC, and glycoproteins linking the ECM and cell membranes, such as VCAN and THBS1. The group also included an unannotated gene, KIAA1199, which may predict to play a role in ECM metabolism based on its association with this cluster. Indeed, this gene has recently been identified as an enzyme involved in the regulation of hyaluronan, a major sugar component of the ECM.

Group 6 contained many genes that encode vesicle transport proteins and proteins associated with cell motility (Figs. 7D and 7E). The vesicle transport genes included microtubule motors and related genes DYNC1H, CKAP1, and factors associated with vesicle formation and trafficking, like DNAJC13 and RAB3B. Again, an unannotated gene, KIAA1462, was found within this cluster. Based on its strong correlation with DYNC1H1 and DNAJC13, it is predicted that this gene may be involved in vesicle transport. The cell motility genes in this group included actin-binding proteins like AFAP1, SPTAN1, SPTBN1, and MYH10, and genes involved in the formation of adhesion complexes, like FLNA and FLNC. Several GTPase-associated factors involved in the regulation of cell motility, attachment and contraction also fell into this group, including DOCK7, ROCK2, IQGAP1,

PRKCA, and AMOTL1. The observation that some cell motility genes correlated with vesicle transport genes is consistent with the role of vesicle transport in cell migration. An additional interesting feature of group 6 is that a subset of these genes, in particular those related to cell motility, were anti-correlated with members of the ECM group discussed above (Fig. 7D). This anti-correlation may reflect regulatory interactions that mediate switching of cells between adherent and migratory states.

Fig. 7 shows cell-to-cell variations and pairwise correlations for the RNA species determined from the 140 gene measurements. Fig. 7A shows a comparison of gene expression levels in two individual cells. Fig. 7B shows Fano factors for individual genes. Error bars represent standard error of the mean determined from 7 independent data sets. Fig. 7C shows Z-scores of the expression variations of four example pairs of genes showing correlated (top two) or anti-correlated (bottom two) variation for 100 randomly selected cells. The Z-score is defined as the difference from the mean normalized by the standard deviation. Fig. 7D is a matrix of the pairwise correlation coefficients of the cell-to-cell variation in expression for the measured genes, shown together with the hierarchical clustering tree. The seven groups identified by a specific threshold on the cluster tree (dashed line) are indicated by the black boxes in the matrix and lines on the tree, with grey lines on the tree indicating ungrouped genes. Different threshold choices on the cluster tree could be made to select either smaller subgroups with tighter correlations or larger super-groups containing more weakly coupled subgroups. Two of the seven groups are enlarged on the right. Fig. 7E shows enrichment of 30 selected, statistically significantly enriched GO terms in the seven groups. Enrichment refers to the ratio of the fraction of genes within a group that have the specific GO term to the fraction of all measured genes having that term. Not all of the GO terms presented here are in the top 10 list.

EXAMPLE 7

This example illustrates mapping spatial distributions of RNAs. As an imaging based approach, MERFISH also allowed the investigation of the spatial distributions of many RNA species simultaneously. Several patterns emerged from the visual inspection of individual genes, with some RNA transcripts enriched in the perinuclear region, some enriched in the cell periphery, and some scattered throughout the cell (Fig. 8A). To identify genes with similar spatial distributions, the correlation coefficients for the spatial density profiles of all pairs of RNA species were determined, and these RNAs were organized based on the pairwise correlations again using the hierarchical clustering approach. The correlation coefficient matrix showed groups of genes with correlated spatial organizations, and the two

most notable groups with the strongest correlations are indicated in Fig. 8B. Group I RNAs appeared enriched in the perinuclear region whereas group II RNAs appeared enriched near the cell peripheral region (Fig. 8C). Quantitative analysis of the distances between each RNA molecule and the cell nucleus or the cell periphery indeed confirmed this visual impression (Fig. 8D).

Group I contained genes encoding extracellular proteins such as FBN1, FBN2 and THSB1, secreted proteins such as PAPPA, and integral membrane proteins such as LRP1 and GPR107. These proteins have no obvious commonalities in function. Rather a GO analysis showed significant enrichment for location terms, such as extracellular region, basement membrane, or perivitelline space (Fig. 8E). To reach these locations, proteins must pass through the secretion pathway, which often requires translation of mRNA at the endoplasmic reticulum (ER). Thus, it is proposed that the spatial pattern that were observed for these mRNAs reflects their co-translational enrichment at the ER. The enrichment of these mRNAs in the perinuclear region (Figs. 8C and 8D, lighter shading), where the rough ER resides, supports this conclusion.

Group II contained genes encoding the actin-binding proteins, including filamins FLNA and FLNC, talin TLN1, and spectrins SPTAN1 and SPTBN1; the microtubule-binding protein CKAP5; and the motor proteins MYH10 and DYNC1H1. This group was enriched with GO terms such as cortical actin cytoskeleton, actin filament binding, and cell-cell adherens junction (Fig. 8E). Beta-actin mRNA may be enriched near the cell periphery in fibroblasts as are mRNAs that encode members of the actin-binding Arp2/3 complex. The enrichment of group II mRNAs in the peripheral region of the cells (Figs. 8C and 8D) suggests that the spatial distribution of the Group II genes might be related to the distribution of actin cytoskeleton mRNAs.

Fig. 8 shows distinct spatial distributions of RNAs observed in the 140-gene measurements. Fig. 8A shows examples of the spatial distributions observed for four different RNA species in a cell. Fig. 8B is a matrix of the pairwise correlation coefficients describing the degree with which the spatial distributions of each gene pair is correlated, shown together with the hierarchical clustering tree. Two strongly correlating groups are indicated by the black boxes on the matrix and shading on the tree. Fig. 8C shows the spatial distributions of all RNAs in the two groups in two example cells. Lighter symbols: group I genes; darker symbols: group II genes. Fig. 8D shows average distances for genes in group I and genes in group II to the cell edge or the nucleus normalized to the average distances for

all genes. Error bars represent SEM across 7 data sets. Fig. 8E shows enrichment of GO terms in each of the two groups.

EXAMPLE 8

This example illustrates measuring 1001 genes with a 14-bit MHD2 code. This example further increases the throughput of MERFISH measurements by simultaneously imaging ~1000 RNA species. This increase could be achieved with the MHD4 code by increasing the number of bits per code word to 32 while maintaining the number of '1' bits per word at four (Fig. 5B). While the stability of the samples across many hybridization rounds (Fig. 14) suggests that such an extension is potentially feasible, an alternative approach is shown here that did not require an increase in the number of hybridizations by relaxing the error correction requirement but keeping the error detection capability. For example, by reducing the Hamming distance from 4 to 2, all 14-bit words could be used that contain four '1' bits to encode 1001 genes and these RNAs were probed with only 14 rounds of hybridization. However, because a single error can produce a word equally close to two different code words, error correction is no longer possible for this modified Hamming-distance-2 (MHD2) code. Hence, it was expected that the calling rate would be lower and the misidentification rate to be higher with this encoding scheme.

To evaluate the performance of this 14-bit MHD2 code, 16 of the 1001 possible code words were set aside as misidentification controls and used the remaining 985 words to encode cellular RNAs. Among these 985 RNAs included 107 RNA species probed in the 140-gene experiments as an additional control. The 1001-gene experiments were performed in IMR90 cells using a similar procedure as described above. To allow all encoding probes to be synthesized from a single 100,000-member oligopool, the number of encoding probes per RNA species was reduced to ~94. Fluorescent spots corresponding to individual RNA molecules were again detected in each round of hybridization with the readout probes and, based on their on-off patterns, these spots were decoded into RNA (Figs. 9A, 19A and 19B). 430 RNA species were detected in the cell shown in Fig. 9A, and similar results were obtained in ~200 imaged cells in 3 independent experiments.

As expected, the misidentification rate of this scheme was higher than that of the MHD4 code. 77% of all real RNA words were detected more frequently than the median count for the misidentification controls instead of the 95% value observed in the MHD4 measurements. Using the same confidence ratio analysis as described above, it was found that 73% (instead of 91% for the MHD4 measurements) of the 985 RNA species were measured with a confidence ratio larger than the maximum value observed for the

misidentification controls (Fig. 19C). RNA copy numbers measured from these 73% RNA species showed excellent correlation with the bulk RNA sequencing results (Pearson correlation coefficient r = 0.76; Fig. 9B, black). It is worth noting that the remaining 27% of the genes still exhibit good, albeit lower, correlation with the bulk RNA sequencing data (r = 0.65; Fig. 9B, red), but the conservative measure of excluding them from further analysis was taken.

The lack of an error correction capability also decreased the calling rate of each RNA species: When comparing the 107 RNA species common in both the 1001-gene and 140-gene measurements, it was found that the copy numbers per cell of these RNA species were lower in the 1001-gene measurements (Figs. 9C and 19D). The total count of these RNAs per cell was ~1/3 of that observed in the 140-gene measurements. Thus the lack of error correction in the MHD2 code produced a ~3-fold decrease in the calling rate, which is consistent with the ~4-fold decrease in calling rate observed for the MHD4 code when error correction was not applied. As expected from the quantitative agreement between 140-gene measurements and conventional smFISH results, comparison of the 1001-gene measurements with conventional smFISH results for 10 RNA species also indicated a ~3-fold drop in calling rate (Fig. 18C). Despite the expected reduction in calling rate, the good correlations found between the copy numbers observed in the 1001-gene measurements and those observed in the 140-gene measurements, as well as in conventional smFISH and bulk RNA sequencing measurements, indicates that the relative abundance of these RNAs can be quantified with the MHD2 encoding scheme.

Simultaneously imaging ~1000 genes in individual cells substantially expanded the ability to detect co-regulated genes. Fig. 10A shows the matrix of pairwise correlation coefficients determined from the cell-to-cell variations in the expression levels of these genes. Using the same hierarchical clustering analysis as described above, ~100 groups of genes with correlated expression were identified. Remarkably, nearly all of these ~100 groups showed statistically significant enrichment of functionally related GO terms (Fig. 10B). These included some of the groups identified in the 140-gene measurements, such as the group associated with cell replication genes and the group associated with cell motility genes (Figs. 10A and 10B, groups 7 and 102), as well as many new groups. The groups identified here included 46 RNA species lacking any previous GO annotations, for which function based on their group association may be hypothesized. For example, KIAA1462 is part of the cell motility group, as also shown in the 140-gene experiments, suggesting a potential role of this gene in cell motility (Fig. 10A, group 102). Likewise, KIAA0355 is part

of a new group enriched in genes associated with heart development (Fig. 10A, group 79), and C17orf70 is part of a group associated with ribosomal RNA processing (Fig. 10A, group 22). Using these groupings, cellular functions for 61 transcription factors and other partially annotated proteins of unknown functions may be hypothesized. For example, the transcription factors Z3CH13 and CHD8 are both members of the cell motility group, suggesting their potential role in the transcriptional regulation of cell motility genes.

Fig. 9 shows simultaneous measurements of 1001 genes in single cells using MERFISH with a 14-bit MHD2 code. Fig. 9A shows the localizations of all detected single molecules in a cell colored based on their measured binary words. Inset: the composite, false-colored fluorescent image of the 14 hybridization rounds for the boxed sub-region with numbered circles indicating potential RNA molecules. Circles indicate unidentifiable molecules, the binary words of which do not match any of the 14-bit MHD2 code words. Images of individual hybridization round are shown in Fig. 19A. Fig. 9B is a scatter plot of the average copy number per cell measured in the 1001-gene experiments versus the abundance measured via bulk sequencing. The upper symbols are for the 73% of genes detected with confidence ratios higher than the maximum ratio observed for the misidentification controls. The Pearson correlation coefficient is 0.76 with a p-value of $3x10^{-133}$. The lower symbols are for the remaining 27% of genes. The Pearson correlation coefficient is 0.65 with a p-value of $3x10^{-33}$. Fig. 9C is a scatter plot of the average copy number for the 107 genes shared in both the 1001-gene measurement with the MHD2 code and the 140-gene measurement with the MHD4 code. The Pearson correlation coefficient is 0.89 with a p-value of $9x10^{-30}$. The dashed line corresponds to the $y = x$ line.

Fig. 10 shows co-variation analysis of the RNA species measured in the 1001-gene measurements. Fig. 10A is a matrix of all pairwise correlation coefficients of the cell-to-cell variation in expression for the measured genes shown with the hierarchical clustering tree. The ~100 identified groups of correlated genes are indicated by shading on the tree. Zoom in of four of the groups described in the text are shown on the right. Fig. 10B is an enrichment of 20 selected, statistically significantly enriched GO terms in the four groups.

Fig. 19 shows decoding and error assessment of the 1001-gene experiments. Fig. 19A shows images of the boxed sub-region of the cell in Fig. 9A for each of 14 hybridization rounds. The final panel is a composite image of these 14 rounds. Circles indicate fluorescent spots that have been identified as potential RNA molecules. Some circles in the composite image indicate unidentifiable molecules, the binary words of which do not match any of the 14-bit MHD2 code words. Fig. 9B shows the corresponding binary word for each of the

spots identified in Fig. 9A with the RNA species to which it is decoded. 'unidentified' implies that the measured binary word does not match any of the 1001 code words. Fig. 19C shows the normalized confidence ratios measured for the 985 RNA species (left) and the 16 misidentification control words not targeted to any RNA (right). The normalized confidence ratio is defined as in Fig. 6F. Fig. 19D shows a histogram of the reduction in detected abundance of 107 genes present in both the 1001-gene experiments and the 140-gene experiments. "Fold decrease in copy number" is defined as the average number of RNA molecules per cell for each species measured in the 140-gene experiments divided by the corresponding average number measured in the 1001-gene experiments.

Fig. 18C is a comparison of the average RNA copy numbers per cell measured in the 1001-gene experiments using the MHD2 code to those determined by conventional smFISH for 10 genes. The average ratio of the copy number measured using the MHD2 measurements to that measured using conventional smFISH was 0.30 +/- 0.05 (mean +/- SEM across 10 genes). The dashed line corresponds to the $y = x$ line and the dotted line corresponds to the $y = 0.30x$ line.

EXAMPLE 9

The above examples illustrate a highly multiplexed detection scheme for systems-level RNA imaging in single cells. Using combinatorial labeling, sequential hybridization and imaging, and two different error-robust encoding schemes, either 140 or 1001 genes in hundreds of individual human fibroblast cells were simultaneously imaged. Of the two encoding schemes presented here, the MHD4 code is capable of both error detection and error correction, and hence can provide a higher calling rate and a lower misidentification rate than the MHD2 code, which instead can only detect but cannot correct errors. MHD2, on the other hand, provides a faster scaling of the degree of multiplexing with the number of bits than MHD4. Other error-robust encoding schemes can also be used for such multiplexed imaging, and experimenters can set the balance between detection accuracy and ease of multiplexing based on the specific requirements of the experiments.

By increasing the number of bits in the code words, it should be possible to further increase the number of detectable RNA species using MERFISH with, for example, a MHD4 or MHD2 code. For example, using the MHD4 code with 32 total bits and four or six '1' bits would increase the number of addressable RNA species to 1,240 or 27,776, respectively. The latter is the approximate scale of the human transcriptome. The predicted misidentification and calling rates are still reasonable for the 32-bit MHD4 code (shown in Figs. 5C and 5D for the MHD4 code with four '1' bits and similar rates were calculated for the MHD4 code with

six '1' bits). If more accurate measurements are desired, an additional increase in the number of bits would allow the use of encoding schemes with a Hamming distance greater than 4, further enhancing the error detection and correction capability. While an increase in the number of bits by adding more hybridization rounds would increase the data collection time and potentially lead to sample degradation, these problems could be mitigated by utilizing multiple colors to readout multiple bits in each round of hybridization.

As the degree of multiplexing is increased, it is important to consider the potential increase in the density of RNAs that need to be resolved in each round of imaging. Based on the imaging and sequencing results, it can be estimated that including the whole transcriptome of the IMR90 cells would lead to a total RNA density of ~200 molecules/micrometer^3. Using the current imaging and analysis methods, 2-3 molecules/ micrometer^3 per hybridization round could be resolved, which would reach a total RNA density of ~20 molecules/micrometer^3 after 32 rounds of hybridization. This density should allow all but the top 10% most expressed genes to be imaged simultaneously or a subset of genes with even higher expression levels to be included. By utilizing more advanced image analysis algorithms to better resolve overlapping images of individual molecules, such as compressed sensing, it is possible to extend the resolvable density by ~4-fold and thus allow all but the top 2% most expressed genes to be imaged all together.

These examples have illustrated the utility of the data derived from highly multiplexed RNA imaging by using co-variation and correlation analysis to reveal distinct sub-cellular distribution patterns of RNAs, to constrain gene regulatory networks, and to predict functions for many previously unannotated or partially annotated genes with unknown functions. Given its ability to quantify RNAs across a wide range of abundances without amplification bias while preserving native context, systems and methods such as MERFISH will allow many applications of *in situ* transcriptomic analyses of individual cells in culture or complex tissues.

EXAMPLE 10

Following are various materials and methods used in the above examples.

Probe design. Each RNA species in the target set was randomly assigned a binary code word either from all 140 possible code words of the 16-bit MHD4 code or from all 1001 possible code words of the 14-bit MHD2 code.

Array-synthesized oligopools were used as templates to make the encoding probes. The template molecule for each encoding probe contained three components: i) a central targeting sequence for *in situ* hybridization to the target RNA, ii) two flanking readout

sequences designed to hybridize each of two distinct readout probes, and iii) two flanking primer sequences to allow enzymatic amplification of the probes (Fig. 13). The readout sequences were taken from the 16 possible readout sequences, each corresponding to one hybridization round. The readout sequences were assigned to the encoding probes such that for any RNA species each of the 4 readout sequences were distributed uniformly along the length of the target RNA and appeared at the same frequency. Template molecules for the 140-gene library also included a common 20-nucleotide (nt) priming region between the first PCR primer and the first readout sequence. This priming sequence was used for the reverse transcription step described below.

Multiple experiments were embedded in a single array-synthesized oligopool, and PCR was used to selectively amplify only the oligos required for a specific experiment. Primer sequences for this indexed PCR reaction were generated from a set of orthogonal 25-nt sequences. These sequences were trimmed to 20 nt and selected for i) a narrow melting temperature range (70 °C to 80 °C), ii) the absence of consecutive repeats of 3 or more identical nucleotides, and iii) the presence of a GC clamp, i.e. one of the two 3' terminal bases must be G or C. To further improve specificity, these sequences were then screened against the human transcriptome using BLAST+, and primers with 14 or more contiguous bases of homology were eliminated. Finally, BLAST+ was again used to identify and exclude primers that had an 11-nt homology region at the 3' end of any other primer or a 5-nt homology region at the 3' end of the T7 promoter. The forward primer sequences (Primer 1) were determined as described above, whereas the reverse primers each contain a 20-nt sequence as described above plus a 20-nt T7 promoter sequence to facilitate amplification via *in vitro* transcription (Primer 2). The primer sequences used in the 140-gene and 1001-gene experiments are listed below.

Table 2

Experiment Name	Primer 1 Sequence (Index Primer 1)	Primer 2 Sequence (T7 promoter plus the reverse complement of Index Primer 2)
140-gene Codebook 1	GTTGGTCGGCACTTGGGTGC (SEQ ID NO: 18)	TAATACGACTCACTATAGGGAAAGCCGGTTCATCCGGTGG (SEQ ID NO: 21)
140-gene Codebook 2	CGATGCGCCAATTCCGGTTC (SEQ ID NO: 19)	TAATACGACTCACTATAGGGTGATCATCGCTCGCGGGTTG (SEQ ID NO: 22)
1001-gene	CGCGGGCTATATGCGAACC	TAATACGACTCACTATAGGGCGTGGAGG

G GCATACAACGC (SEQ ID NO: 23)
(SEQ ID NO: 20)

30-nt-long readout sequences were created by concatenating fragments of the same orthogonal primer set generated above by combining one 20-nt primer with a 10-nt fragment of another. These readout sequences were then screened, using BLAST+, for orthogonality with the index primer sequences and other readout sequences (no more than 11 nt of homology) and for potential off-target binding sites in the human genome (no more than 14 nt of homology). Fluorescently labeled readout probes with sequences complementary to the readout sequences were used to probe these readout sequences, one in each hybridization round. All used readout probes sequences are listed below.

Table 3

Bit	Readout probes	
1	CGCAACGCTTGGGACGGTTCCAATCGGATC/3Cy5Sp/	SEQ ID NO: 24
2	CGAATGCTCTGGCCTCGAACGAACGATAGC/3Cy5Sp/	SEQ ID NO: 25
3	ACAAATCCGACCAGATCGGACGATCATGGG/3Cy5Sp/	SEQ ID NO: 26
4	CAAGTATGCAGCGCGATTGACCGTCTCGTT/3Cy5Sp/	SEQ ID NO: 27
5	GCGGGAAGCACGTGGATTAGGGCATCGACC/3Cy5Sp/	SEQ ID NO: 28
6	AAGTCGTACGCCGATGCGCAGCAATTCACT/3Cy5Sp/	SEQ ID NO: 29
7	CGAAACATCGGCCACGGTCCCGTTGAACTT/3Cy5Sp/	SEQ ID NO: 30
8	ACGAATCCACCGTCCAGCGCGTCAAACAGA/3Cy5Sp/	SEQ ID NO: 31
9	CGCGAAATCCCCGTAACGAGCGTCCCTTGC/3Cy5Sp/	SEQ ID NO: 32
10	GCATGAGTTGCCTGGCGTTGCGACGACTAA/3Cy5Sp/	SEQ ID NO: 33
11	CCGTCGTCTCCGGTCCACCGTTGCGCTTAC/3Cy5Sp/	SEQ ID NO: 34
12	GGCCAATGGCCCAGGTCCGTCACGCAATTT/3Cy5Sp/	SEQ ID NO: 35
13	TTGATCGAATCGGAGCGTAGCGGAATCTGC/3Cy5Sp/	SEQ ID NO: 36
14	CGCGCGGATCCGCTTGTCGGGAACGGATAC/3Cy5Sp/	SEQ ID NO: 37
15	GCCTCGATTACGACGGATGTAATTCGGCCG/3Cy5Sp/	SEQ ID NO: 38
16	GCCCGTATTCCCGCTTGCGAGTAGGGCAAT/3Cy5Sp/	SEQ ID NO: 39

The readout probes used for the 140-gene libraries were probes 1 through 16. The readout probes used for the 1001-gene experiment were probes 1 through 14. /3Cy5Sp/ indicates a 3' Cy5 modification.

To design the central targeting sequences of the encoding probes, the abundance of different transcripts in IMR90 cells using Cufflinks v2.1, total RNA data from the ENCODE project, and human genome annotations from Gencode v18 were complied. Probes were designed from gene models corresponding to the most abundant isoform using OligoArray2.1 with the following constraints: the target sequence region is 30-nt long; the melting temperatures of the hybridized region of the probe and cellular RNA target is greater than 70 °C; there is no cross hybridization targets with melting temperatures greater than 72 °C; there

is no predicted internal secondary structures with melting temperatures greater than 76 °C; and there is no contiguous repeats of 6 or more identical nucleotides. Melting temperatures were adjusted to optimize the specificity of these probes and minimize secondary structure while still producing sufficient numbers of probes for the libraries. To decrease computational cost, isoforms were divided into 1-kb regions for probe design. Using BLAST+, all potential probes that mapped to more than one cellular RNA species were rejected. Probes with multiple targets on the same RNA were kept.

For each gene in the 140-gene experiments, 198 putative encoding probe sequences were generated by concatenating the appropriate index primers, readout sequences, and targeting regions as shown in Fig. 13. To address the possibility that concatenation of these sequences introduced new regions of homology to off-target RNAs, BLAST+ was used to screen these putative sequences against all human rRNA and tRNA sequences as well as highly expressed genes (genes with FPKM > 10,000). Probes with greater than 14 nt of homology to rRNAs or tRNAs or greater than 17 nt of homology to highly expressed genes were removed. After these cuts, there were ~192 (with a standard deviation of 2) probes per gene for both MHD4 codebooks used in the 140-gene experiments. The same protocol for the 1001-gene experiments was used, as follows: Starting with 96 putative targeting sequences per gene, ~94 (with a standard deviation of 6) encoding probes per gene were obtained after these additional homology cuts. The number of encoding probes per RNA was decreased for the 1001-gene experiments so that these probes could be synthesized from a single 100,000-member oligopool as opposed to two separate pools. Each encoding probe was designed to contain two of the four readout sequences associated with each code word, hence only half of the bound encoding probes can bind readout probe during any given hybridization round. ~192 or ~94 encoding probes per RNA were used to obtain high signal-to-background ratios for individual RNA molecules. The number of encoding probes per RNA could be substantially reduced but still allow single RNA molecules to be identified. In addition, increasing the number of readout sequences per encoding probe or using optical sectioning methods to reduce the fluorescence background may allow further reduction in the number of the encoding probes per RNA.

Two types of misidentification controls were designed. The first control (blank words) were not represented with encoding probes. The second type of control (no-target words) had encoding probes that were not targeting any cellular RNA. The targeting regions of these probes were composed of random nucleotide sequences subject to the same constraints used to design the RNA targeting sequences described above. Moreover, these

random sequences were screened against the human transcriptome to ensure that they contain no significant homology (>14-nt) to any human RNA. The 140-gene measurements contained 5 blank words and 5 no-target words. The 1001-gene measurements contained 11 blank words and 5 no-target words.

Probe synthesis. The encoding probes were synthesized using the following steps, and this synthesis protocol is illustrated in Fig. 13.

Step 1: The template oligopool (CustomArray) was amplified via limited-cycle PCR on a Bio-Rad CFX96 using primer sequences specific to the desired probe set. To facilitate subsequent amplification via *in vitro* transcription, the reverse primer contained the T7 promoter. All primers were synthesized by IDT. This reaction was column purified (Zymo DNA Clean and Concentrator; D4003).

Step 2: The purified PCR products were then further amplified ~200-fold and converted into RNA via a high yield *in vitro* transcription according to the manufacturer's instructions (New England Biolabs, E2040S). Each 20 microliter reaction contained ~1 microgram of template DNA from above, 10 mM of each NTP, 1x reaction buffer, 1x RNase inhibitor (Promega RNasin, N2611) and 2 microliters of the T7 polymerase. This reaction was incubated at 37 °C for 4 hours to maximize yield. This reaction was not purified before the following steps.

Step 3: The RNA products from the above *in vitro* transcription reaction were then converted back into DNA via a reverse transcription reaction. Each 50 microliter reaction contained the unpurified RNA produce from Step 2 supplemented with 1.6 mM of each dNTP, 2 nmol of a reverse transcription primer, 300 units of Maxima H- reverse transcriptase (Thermo Scientific, EP0751), 60 units of RNasin, and a final 1x concentration of the Maxima RT buffer. This reaction was incubated at 50 °C for 45 minutes, and the reverse transcriptase was inactivated at 85 °C for 5 minutes. The templates for the 140-gene libraries contain a common priming region for this reverse transcription step; thus, a single primer was used for this step when creating these probes. Its sequence was CGGGTTTAGCGCCGGAAATG (SEQ ID NO: 40). A common priming region was not included for the 1001-gene library; thus, the reverse transcription was conducted with the forward primer: CGCGGGCTATATGCGAACCG (SEQ ID NO: 20).

Step 4: To remove the template RNA, 20 microliters of 0.25 M EDTA and 0.5 N NaOH was added to the above reaction to selectively hydrolyze RNA, and the sample was incubated at 95 °C for 10 minutes. This reaction was then immediately purified by column purification using a 100-microgram-capacity column (Zymo Research, D4030) and the Zymo

Oligo Clean and Concentrator protocol. The final probes were eluted in 100 microliters of RNase-free deionized water, evaporated in a vacuum concentrator, and then resuspended in 10 μL of encoding hybridization buffer (see below). Probes were stored at -20 °C. Denaturing poly-acrylamide gel electrophoresis and absorption spectroscopy were used to confirm the quality of the probes and revealed that this probe synthesis protocol converts 90-100% of the reverse-transcription primer into full length probe and of the probe that is constructed, 70-80% is recovered during the purification step.

Fluorescently labeled readout probes have sequences complementary to the readout sequences described above and a Cy5 dye attached at the 3' end. These probes were obtained from IDT and HPLC purified.

Sample preparation and labeling with encoding probes. Human primary fibroblasts (American Type Culture Collection, IMR90) were used in this work. These cells are relatively large and flat, facilitating wide-field imaging without the need for optical sectioning. Cells were cultured with Eagle's Minimum Essential Medium. Cells were plated on 22-mm, #1.5 coverslips (Bioptechs, 0420-0323-2) at 350,000 cells/coverslip and incubated at 37 °C with 5% CO_2 for 48-96 hours within petri dishes. Cells were fixed for 20 minutes in 4% paraformaldehyde (Electron Microscopy Sciences, 15714) in 1x phosphate buffered saline (PBS; Ambion, AM9625) at room temperature, reduced for 5 minutes with 0.1% w/v sodium borohydride (Sigma, 480886) in water to reduce background fluorescence, washed three times with ice-cold 1x PBS, permeabilized for 2 minutes with 0.5% v/v Triton (Sigma, T8787) in 1x PBS at room temperature, and washed three times with ice cold 1x PBS.

Cells were incubated for 5 minutes in encoding wash buffer comprising 2x saline-sodium citrate buffer (SSC) (Ambion, AM9763), 30% v/v formamide (Ambion, AM9342), and 2 mM vanadyl ribonucleoside complex (NEB, S1402S). 10 microliters of 100 micromolar(140-gene experiments) or 200 micromolar (1001-gene experiments) encoding probes in encoding hybridization buffer was added to the cell-containing coverslip and spread uniformly by placing another coverslip on top of the sample. Samples were then incubated in a humid chamber inside a 37 °C-hybridization oven for 18-36 hours. Encoding hybridization buffer is composed of encoding wash buffer supplemented with 1 mg/mL yeast tRNA (Life Technologies, 15401-011) and 10% w/v dextran sulfate (Sigma, D8906-50G).

Cells were then washed with primary encoding wash buffer, incubated at 47 °C for 10 minutes, and this wash was repeated for a total of three times. A 1:1000 dilution of 0.2-micrometer-diameter carboxylate-modified orange fluorescent beads (Life Technologies, F-8809) in 2xSSC was sonicated for 3 minutes and then incubated with the sample for 5

minutes. The beads were used as fiducial markers to align images obtained from multiple successive rounds of hybridization, as described below. The sample was washed once with 2xSSC, and then post-fixed with 4% v/v paraformaldehyde in 2xSSC at room temperature for 30 minutes. The sample was then washed three times with 2xSSC and either imaged immediately or stored for no longer than 12 hours at 4 °C prior to imaging. All solutions were prepared as RNase-free.

MERFISH imaging with multiple successive rounds of hybridization. The sample coverslip was assembled into a Bioptech's FCS2 flow chamber, and the flow through this chamber was controlled via a home-built fluidics system composed of three computer-controlled 8-way valves (Hamilton, MVP and HVXM 8-5) and a computer-controlled peristaltic pump (Rainin, Dynamax RP-1). The sample was imaged on a home-built microscope constructed around an Olympus IX-71 body and a 1.45 NA, 100x oil immersion objective and configured for oblique incidence excitation. The objective was heated to 37 °C with a Bioptechs objective heater. Constant focus was maintained throughout the imaging process with a home-built, auto-focusing system. Illumination was provided at 641 nm, 561 nm, and 405 nm using solid state lasers (MPB communications, VFL-P500-642; Coherent, 561-200CWCDRH; and Coherent, 1069413/AT) for excitation of the Cy5-labeled readout probes, the fiducial beads, and nuclear counterstains, respectively. These lines were combined with a custom dichroic (Chroma, zy405/488/561/647/752RP-UF1) and the emission was filtered with a custom dichroic (Chroma, ZET405/488/561/647-656/752m). Fluorescence was separated with a QuadView (Photometrics) using the dichroics T560lpxr, T650lpxr, 750dcxxr (Chroma) and the emission filters ET525/50m, WT59550m-2f, ET700/75m, HQ770lp (Chroma) and imaged with an EMCCD camera (Andor, iXon-897). The camera was configured so that a pixel corresponds to 167 nm in the sample plane. The entire system was fully automated, so that imaging and fluid handling were performed for the entire experiment without user intervention.

Sequential hybridization, imaging, and bleaching proceeded as follows. 1 mL of 10 nM of the appropriate fluorescently labeled readout probe in readout hybridization buffer (2xSSC; 10% v/v formamide; 10% w/v dextran sulfate, and 2 mM vanadyl ribonucleoside complex) was flown across the sample, flow was stopped, and the sample was incubated for 15 minutes. Then 2 mL of readout wash buffer (2xSSC, 20% v/v formamide; and 2 mM vanadyl ribonucleoside complex) was flown across the sample, flow was stopped, and the sample was incubated for 3 minutes. 2 mL of imaging buffer comprising 2xSSC, 50 mM TrisHCl pH 8, 10% w/v glucose, 2 mM Trolox (Sigma-Aldrich, 238813), 0.5 mg/mL glucose

oxidase (Sigma-Aldrich, G2133), and 40 microgram/mL catalase (Sigma-Aldrich, C30) was flown across the sample. Flow was then stopped, and then approximately 75 to 100 regions were exposed to ~25 mW 642-nm and 1 mW of 561-nm light and imaged. Each region was 40 micrometers by 40 micrometers. The laser powers were measured at the microscope backport. Because the imaging buffer is sensitive to oxygen, the ~50 mL of imaging buffer used for a single experiment was made fresh at the beginning of the experiment and then stored under a layer of mineral oil throughout the measurement. Buffer stored in this fashion was stable for more than 24 hours.

After imaging, the fluorescence of the readout probes was extinguished via photobleaching. The sample was washed with 2 mL of photobleaching buffer (2xSSC and 2 mM vanadyl ribonucleoside complex), and each imaged region of the sample was exposed to 200 mW of 641-nm light for 3 s. To confirm the efficacy of this photobleaching treatment, imaging buffer was reintroduced, and the sample was imaged as described above.

The above hybridization, imaging, and photobleaching process was repeated either 16 times for the 140-gene measurements using the MHD4 code or 14 times for the 1001-gene measurements using the MHD2 code. An entire experiment was typically completed in ~20 hours.

Following completion of imaging, 2 mL of a 1:1000 dilution of Hoescht (ENZ-52401) in 2xSSC was flown through the chamber to label the nuclei of the cells. The sample was then washed immediately with 2 mL of 2xSSC followed by 2 mL of imaging buffer. Each region of the sample was then imaged once again with ~1 mW of 405-nm light.

Because cells were imaged using wide-field imaging with oblique-incidence illumination, without optical sectioning and z-scanning, the fraction of individual RNA species that was outside the axial range of the imaging geometry was quantified for 6 different RNA species using conventional smFISH. For this purpose, these cells were optically sectioned by collecting stacks of images at different focal depths through the entire depth of the cells. The images were aligned in consecutive focal planes and then computed for each cell the fraction of RNAs that were detected in the three-dimensional stack but not in the basal focal plane. It was found that only a small fraction, 15% +/- 1% (Mean +/- SEM across six different RNA species) of RNA molecules were outside the imaging range of a fixed focal plane without z-scanning. These measurements also confirmed that the excitation geometry illuminated the full depth of the cells. Any optical sectioning technique could be employed in MERFISH to allow the imaging of RNAs in thicker cells or tissues.

Construction of measured words. Fluorescent spots were identified and localized in each image using a multi-Gaussian-fitting algorithm assuming a Gaussian with a uniform width of 167 nm. This algorithm was used to allow partially overlapping spots to be distinguished and individually fit. RNA spots were distinguished from background signal, i.e. signal arising from probes bound non-specifically, by setting the intensity threshold required to fit a spot with this software. Due to variation in the brightness of spots between rounds of hybridization, this threshold was adjusted appropriately for each hybridization round to minimize the combined average of the 1-->0 and 0-->1 error rates across all hybridization rounds (140-gene measurements) or to maximize the ratio of the number of measured words with four '1' bits to those with three or five '1' bits (1001-gene measurements). The location of the fiducial beads was identified in each frame using a faster single-Gaussian fitting algorithm.

Images of the same sample region in different rounds of hybridization were registered by rotating and translating the image to align the two fiducial beads within the same image that were most similar in location after a coarse initial alignment via image correlation. All images were aligned to a coordinate system established by the images collected in the first round of hybridization. The quality of this alignment was determined from the residual distance between five additional fiducial beads, and alignment error was typically ~20 nm.

Fluorescence spots in different hybridization rounds were connected into a single string, corresponding to a potential RNA molecule, if the distance between spots was smaller than 1 pixel (167 nm). For each string of spots, the on-off sequence of fluorescent signals in all hybridization rounds were used to assign a binary word to the potential RNA molecule, in which '1' was assigned to the hybridization rounds that contained a fluorescent signal above threshold and '0' was assigned to the other hybridization rounds. Measured words were then decoded into RNA species using the 16-bit MHD4 code or the 14-bit MHD2 code discussed above. In the case of the 16-bit MHD4 code, if the measured binary word matched the code word of a specific RNA perfectly or differed from the code word by one single bit, it was assigned to that RNA. In the case of the 14-bit MHD2 code, only if the measured binary word matched the code word of a specific RNA perfectly, was it assigned to that RNA. To determine the copy number per cell, the number of each RNA species was counted in individual cells within each 40 micrometer by 40 micrometer imaging area. It is noted that this number accounts for the majority but not all RNA molecules within a cell because a fraction of the cell could be outside the imaging area or focal depth. Tiling images of

adjacent areas and adjacent focal planes could be employed to improve the counting accuracy.

In the 140-gene experiments, some regions of the cell nucleus occasionally contained too much fluorescence signal to properly identify individual RNA spots. In the 1001-gene experiments, the cell nucleus generally contained too much fluorescent signal to allow identification of individual RNA molecules. These bright regions were excluded from all subsequent analysis. This work focuses on mRNAs, which are enriched in the cytoplasm. To estimate the fraction of mRNAs missed by excluding the nucleus region, conventional smFISH was used to quantify the fraction of molecules found inside the nucleus for six different mRNAs species. It was found that only 5% +/- 2% (Mean +/- SEM across six RNA species) of these RNA molecules are found in the nucleus. Employment of super-resolution imaging and/or optical sectioning could potentially allow individual molecules in these dense nucleus regions to be identified, which would be particularly useful for probing those non-coding RNAs that are enriched in the nucleus.

smFISH measurements of individual genes. Pools of 48 fluorescently-labeled (Quasar 670) oligonucleotide probes per RNA were purchased from Biosearch Technologies. 30-nt probe sequences were taken directly from a random subset of the targeting regions used for the multiplexed measurements. Cells were fixed and permeabilized as described above. 10 microliters of 250 nM oligonucleotide probes in encoding hybridization buffer (described above) was added to the cell-containing coverslip and spread uniformly by placing another coverslip on top of the sample. Samples were then incubated in a humid chamber inside a 37 °C-hybridization oven for 18 hours. Cells were then washed with encoding wash buffer (described above) at 37 °C for 10 minutes, and this wash was repeated for a total of three times. The sample was then washed three times with 2xSSC and imaged in imaging buffer using the same imaging geometry as described above for MERFISH.

Bulk RNA sequencing. Total RNA was extracted from IMR90 cells cultured as above using the Zymo Quick RNA MiniPrep kit (R1054) according to the manufacturer's instructions. polyA RNA was then selected (NEB; E7490), and a sequencing library was constructed using the NEBNext Ultra RNA library preparation kit (NEB; E7530), amplified with custom oligonucleotides, and 150-bp reads were obtained from on a MiSeq. These sequences were aligned to the human genome (Gencode v18) and isoform abundance was computed with cufflinks.

Calculation of the predicted scaling and error properties of different encoding schemes. Analytic expressions were derived for the dependence of the number of possible

code words, the calling rate, and the misidentification rate on *N*. The calling rate is defined as the fraction of RNA molecules that are properly identified. The misidentification rate is defined as the fraction of RNA molecules that are misidentified as a wrong RNA species. For encoding schemes with an error-detection capability, the calling rate and misidentification rate does not add up to 1 because a fraction of the molecules not called properly can be detected as errors and discarded and, hence, not misidentified as a wrong species. These calculations assume that the probability of misreading bits is constant for all hybridization rounds but differs for the 1-->0 and 0-->1 errors. Experimentally measured average 1-->0 and 0-->1 error rates (10% and 4% respectively) were used for the estimates shown in Figs. 5B-5D. For simplicity, the word corresponding to all '0's was not removed from calculations.

For the simple binary encoding scheme in which all possible *N*-bit binary words are assigned to unique RNA species, the number of possible code words is 2^N. The number of words that could be used to encode RNA is actually $2^N - 1$ because the code word '00...0' does not contain detectable fluorescence in any hybridization round, but for simplicity the word corresponding to all '0's was not removed from subsequent calculations. The error introduced by this approximation is negligible. For any given word with *m* '1's and *N*–*m* '0's the probability of measuring that word without error (the fraction of RNAs that is properly called) is:

$$(1-p_1)^m(1-p_0)^{N-m}, \tag{1}$$

where p_1 is 1-->0 error rate and p_0 is 0-->1 error rate per bit. Because different words in this simple binary encoding scheme can have different numbers of '1' bits, the calling rate for different words will differ if $p_1 \neq p_0$. The average calling rate, reported in Fig. 5C, was determined from the weighted average of the value of Eq. (1) for all words. This weighted average is:

$$\frac{1}{2^N}\sum_{m=0}^{N}\binom{N}{m}(1-p_1)^m(1-p_0)^{N-m}, \tag{2}$$

where $\binom{N}{m}$ is the binomial coefficient and corresponds to the number of words with *m* '1' bits in this encoding scheme. Since in this encoding scheme every error produces a binary word that encodes a different RNA, the average misidentification rate for this encoding scheme, reported in Fig. 5D, follows directly from (2):

$$1-\frac{1}{2^N}\sum_{m=0}^{N}\binom{N}{m}(1-p_1)^m(1-p_0)^{N-m}. \quad (3)$$

To calculate the scaling and error properties of the extended Hamming distance 4 (HD4) code, the generator matrix for the desired number of data bits using standard methods was first created. The generator matrix determines the specific words that are present in a given encoding scheme and was used to directly determine the number of encoded words as a function of the number of bits. In this encoding scheme, the calling rate corresponds to the fraction of words measured without error as well as the fraction of words measured with a single-bit error. For code words with m '1' bits, this fraction is determined by the following expression:

$$(1-p_1)^m(1-p_0)^{N-m}+m\,p_1{}^1(1-p_1)^{m-1}(1-p_0)^{N-m}+(N-m)\,p_0{}^1(1-p_1)^m(1-p_0)^{N-m-1} \quad (4)$$

where the first term is the probability of not making any errors, the second term corresponds to the total probability of making one 1-->0 error at any of the m '1' bits without making any other 0-->1 errors, and the final term corresponds to the total probability of making one 0-->1 error at any of the N-m '0' bits without making any 1-->0 errors. Because the number of '1' bits can differ between words in this encoding scheme, the average calling rate reported in Fig. 5C was computed from a weighted average over Eq. (4) for different values of m. The weight for each term was determined from the number of words that contain m '1' bits as determined from the generator matrix described above.

Because RNA-encoding words are separated by a minimum Hamming distance of 4, at least 4 errors are required to switch one word into another. If error correction is applied, then 3 or 5 errors could also convert one RNA into another. Thus, the misidentification rate from all possible combinations of 3-bit, 4-bit and 5-bit errors was estimated for code words with m '1' bits. Technically, >5-bit errors could also convert one RNA into another, but the probability of making such errors is negligible because of the small per-bit error rate. This expression was approximated with:

$$\begin{aligned}
&\sum_{i=0}^{4}\binom{m}{i}\binom{N-m}{4-i}p_1^i p_0^{4-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(4-i)}+\\
&\sum_{i=0}^{3}\binom{m}{i}\binom{N-m}{3-i}p_1^i p_0^{3-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(3-i)}+\\
&\sum_{i=0}^{5}\binom{m}{i}\binom{N-m}{5-i}p_1^i p_0^{5-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(5-i)}+
\end{aligned} \quad (5)$$

The first sum corresponds to all of the ways in which exactly four mistakes can be made. Similarly, the second and third sums correspond to all of the ways in which exactly three or

five mistakes can be made. Eq. (5) provides an upper bound for the misidentification rate because not all three, four, or five bit errors produce a word that matches or would be corrected to another legitimate word. Again because the number of '1' bits can differ between words, the average misidentification rate reported in Fig. 5D is calculated as a weighted average of Eq. (5) over the number of words that have *m* '1' bits.

To generate the MHD4 code where the number of '1' bits for each code word is set to 4, the HD4 codes were first generated as described above, and then all code words that did not contain four '1's were removed. The calling rate of this code, reported in Fig. 5C, was directly calculated from Eq. (4) but with m = 4 because all code words in this code have four '1' bits. The misidentification rate of this code, reported in Fig. 5D, was calculated by modifying Eq. (5) with the following considerations: (i) the number of '1' bits, m, was set to 4 and (ii) errors that produce words that do not contain three, four, or five '1' bits were excluded. Thus, the expression in Eq. (5) was simplified to

$$\begin{aligned}
&\binom{4}{2}\binom{N-4}{2}p_1^2 p_0^2(1-p_1)^2(1-p_0)^{N-6} + \\
&\binom{4}{1}\binom{N-4}{2}p_1^1 p_0^2(1-p_1)^3(1-p_0)^{N-6} + \binom{4}{2}\binom{N-4}{1}p_1^2 p_0^1(1-p_1)^2(1-p_0)^{N-5} \\
&\binom{4}{2}\binom{N-4}{3}p_1^2 p_0^3(1-p_1)^2(1-p_0)^{N-7} + \binom{4}{3}\binom{N-4}{2}p_1^3 p_0^2(1-p_1)^1(1-p_0)^{N-6}
\end{aligned} \tag{6}$$

Again, this expression is an upper bound on the actual misidentification rate because not all words with four '1's are valid code words.

Estimates of the 1-->0 and 0-->1 error rates for each hybridization round. To compute the probability of misreading a bit at a given hybridization round, the error correcting properties of the MHD4 code were used. Briefly, the probabilities of 1-->0 or 0-->1 errors were derived in the following way. Let the probability of making an error at the *i*th bit, i.e. *i*th hybridization round, be p_i and the actual number of RNA molecules of the given species be *A*, then the number of exact matches for this RNA will be $W_E = A\prod_{i=1}^{16}(1-p_i)$ and the number of one-bit error corrected matches for this RNA corresponding to errors at the *i*th bit will be $W_i = A\frac{p_i}{(1-p_i)}\prod_{j=1}^{16}(1-p_j)$. The p_i can be directly derived from the ratio: $W_i/W_E = \frac{p_i}{(1-p_i)}$. This ratio assumes that the one-bit error-corrected counts were only generated from single-bit errors from the correct word and that multi-error contamination from other RNA words is negligible. Given that the error rate per hybridization round is small and that it takes at least

three errors to convert one RNA-encoding word into a word that would be misidentified as another RNA, the above approximation should be a good one.

To compute the average 1-->0 or 0-->1 error probabilities for each of the 16 hybridization rounds, the above approach was used to calculate the per-bit error rates for each bit of every gene, and these errors were sorted based on whether they correspond to a 1-->0 or a 0-->1 error, and the average of these errors for each bit weighted by the number of counts observed for the corresponding gene was taken.

Estimates of the calling rate for individual RNA species from actual imaging data. With the estimates of the 1-->0 or 0-->1 error probabilities for each round of hybridization as determined above, it is possible to estimate the calling rate for each RNA based on the specific word used to encoded it. Specifically, the fraction of an RNA species that is called correctly is determined by

$$\prod_{i=1}^{N}(1-p_i)+\sum_{j=1}^{N}\frac{p_j}{(1-p_j)}\prod_{i=1}^{N}(1-p_i), \tag{7}$$

where the first term represent the probability of observing an exact match of the code word and the second term represent the probability of observing an error-corrected match (i.e. with one-bit error). The values of the per-bit error rate p_i for each RNA species are determined by the specific code word for that RNA and the measured 1-->0 or 0-->1 error rates for each round of hybridization. If the code word of the RNA contains a '1' in the *i*th bit, then p_i is determined from the 1-->0 error rate for the *i*th hybridization round; if the word contains a '0' in the *i*th bit, p_i is determined from the 0-->1 error rate for the *i*th hybridization round.

Hierarchical clustering analysis of the co-variation in RNA abundance. Hierarchical clustering of the co-variation in gene expression for both the 140-gene and 1001-gene experiments was conducted as follows. First, the distance between every pair of genes was determined as 1 minus the Pearson correlation coefficient of the cell-to-cell variation of the measured copy numbers of these two RNA species, both normalized by the total RNA counted in the cell. Thus, highly correlated genes are 'closer' to one another and highly anti-correlated genes are 'further' apart. An agglomerative hierarchical cluster tree was then constructed from these distances using the Unweighted Pair Group Method with Arithmetic mean (UPGMA). Specifically, starting with individual genes, hierarchical clusters were constructed by identifying the two clusters (or individual genes) that are closest to one another according to the arithmetic mean of the distances between all inter-cluster gene pairs. The pairs of clusters (or individual genes) with the smallest distance are then grouped

together and the process is repeated. The matrix of pairwise correlations was then sorted based on the order of the genes within these trees.

Groups of genes with substantial co-variations were identified by selecting a threshold on the hierarchical cluster tree (indicated by the dashed lines in Figs. 7D and 10A) that produced approximately 10 groups of genes each of which contains at least 4 members for the 140-gene experiments or approximately 100 groups each of which contains at least 3 members for the 1001-gene experiments. It is noted that one can change the threshold in order to identify either more tightly coupled smaller groups or larger groups with relatively loose coupling.

A probability value for the confidence that a gene belongs to a specific group was determined by computing the difference between the average correlation coefficient between that gene and all other members of that group and the average correlation coefficient between that gene and all other measured genes outside that group. The significance (p-value) of this difference was determined with the student's t-test.

Because hierarchical clustering is inherently a one-dimensional analysis, i.e. any given genes can only be a member of a single group, this analysis does not allow all correlated gene groups to be identified. Higher dimension analysis, such as principal component analysis or k-means clustering, could be used to identify more co-varying gene clusters.

Analysis of RNA spatial distributions. To identify genes that have similar spatial distributions, each of the measured cells was subdivided into 2x2 regions and calculated the fraction of each RNA species present in each of these bins. To control for the fact that some regions of the cell naturally contain more RNA than others, the enrichment was calculated for each gene , i.e., the ratio of the observed fraction in a given region for a given RNA species to the average fraction observed for all genes in that same region. For each pair of RNA species, the Pearson correlation coefficient of the region-to-region variation in enrichment of these two RNA species for each cell was determined and the correlation coefficients were averaged over ~400 cells imaged in 7 independent data sets. RNA species were then clustered based on these average correlation coefficients using the same hierarchical clustering algorithm described above. Because of the large number of cells used for the analysis, it was found that the coarse spatial binning (2x2 regions per cell) was sufficient to capture the spatial correlation between genes and finer binning did not produce more significantly correlated groups.

To measure the distances of genes from the nuclei and from the cell edge, brightness thresholds on the cell images were first used to segment the nuclei and the cell edges identified. The distance from every RNA molecule to the nearest part of the nucleus and nearest part of the cell edge was then determined. For each data set, the average distance for each RNA species averaged over all the cells measured was then computed. These distances were averaged for the group I genes, group II genes or all genes. Only those RNA species with at least 10 counts per cell were used in this analysis to minimize statistical error on the distance values.

Gene ontology (GO) analysis. Groups of genes were selected from the hierarchical trees as discussed above. A collection of GO terms was determined for all measured RNA species as well as the RNA species associated with each group from the most recent human GO annotations using both the annotated GO terms and terms immediately upstream or downstream of the found annotations. The enrichment of these annotations was calculated from the ratio of the fraction of genes within each group that have this term to the fraction of all measured genes that have this term and the p-value for this enrichment was calculated via the hypergeometric function. Only statistically significantly enriched GO terms with a p-value less than 0.05 were considered.

While several embodiments of the present invention have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the functions and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the present invention. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the teachings of the present invention is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, the invention may be practiced otherwise than as specifically described and claimed. The present invention is directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles,

materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the scope of the present invention.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

The indefinite articles "a" and "an," as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean "at least one."

The phrase "and/or," as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with "and/or" should be construed in the same fashion, i.e., "one or more" of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the "and/or" clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to "A and/or B", when used in conjunction with open-ended language such as "comprising" can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, "or" should be understood to have the same meaning as "and/or" as defined above. For example, when separating items in a list, "or" or "and/or" shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as "only one of" or "exactly one of," or, when used in the claims, "consisting of," will refer to the inclusion of exactly one element of a number or list of elements. In general, the term "or" as used herein shall only be interpreted as indicating exclusive alternatives (i.e. "one or the other but not both") when preceded by terms of exclusivity, such as "either," "one of," "only one of," or "exactly one of." "Consisting essentially of," when used in the claims, shall have its ordinary meaning as used in the field of patent law.

As used herein in the specification and in the claims, the phrase "at least one," in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also

allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase "at least one" refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, "at least one of A and B" (or, equivalently, "at least one of A or B," or, equivalently "at least one of A and/or B") can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

When the word "about" is used herein in reference to a number, it should be understood that still another embodiment of the invention includes that number not modified by the presence of the word "about."

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

In the claims, as well as in the specification above, all transitional phrases such as "comprising," "including," "carrying," "having," "containing," "involving," "holding," "composed of," and the like are to be understood to be open-ended, i.e., to mean including but not limited to. Only the transitional phrases "consisting of" and "consisting essentially of" shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

RELATED APPLICATIONS

This application is a national stage filing of International Patent Application Serial No. PCT/US2015/042556, filed July 29, 2017, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*, which claims the benefit of U.S. Provisional Patent Application Serial No. 62/031,062, filed July 30, 2014, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/142,653, filed April 3, 2015, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; and U.S. Provisional Patent Application Serial No. 62/050,636, filed September 15, 2014, entitled "Probe Library Construction," by Zhuang, *et al.* Each of the above is incorporated herein by reference.

GOVERNMENT FUNDING

This invention was made with government support under Grant No. GM096450 awarded by National Institutes of Health. The government has certain rights in the invention.

FIELD

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined.

BACKGROUND

Single-molecule fluorescent in situ hybridization (smFISH) is a powerful method for detecting individual mRNA molecules in cells. The high detection efficiency and large dynamic range of this method provides exquisite detail into the expression state, spatial distribution within cells and intact tissues, and variation among cells of individual mRNAs. Such approaches have been essential to many recent insights into understanding gene regulation and expression. A fundamental limitation of smFISH, however, is its low throughput, typically only a few genes at a time. This low throughput is due to a lack of distinguishable probes with which to label cells and the cost of producing large amounts of labeled probe required for high efficiency staining. Thus, improvements in detecting mRNA molecules are needed.

SUMMARY

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. The subject matter of the present invention

involves, in some cases, interrelated products, alternative solutions to a particular problem, and/or a plurality of different uses of one or more systems and/or articles.

In one aspect, the present invention is generally directed to a composition. According to one set of embodiments, the composition comprises a plurality of nucleic acid probes, at least some of which comprise a first portion comprising a target sequence and a plurality of read sequences. In some cases, each comprises a first portion comprising a target sequence and a plurality of read sequences. In some embodiments, the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

In another aspect, the present invention is generally directed to a method. In one set of embodiments, the method includes acts of exposing a sample to a plurality of nucleic acid probes; for each of the nucleic acid probes, determining binding of the nucleic acid probes within the sample; creating codewords based on the binding of the nucleic acid probes; and for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

The method, in another set of embodiments, includes acts of exposing a sample to a plurality of nucleic acid probes, wherein the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences, and wherein at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences; and for each of the nucleic acid probes, determining binding of the target sequences of the nucleic acid probes within the sample.

In yet another set of embodiments, the method includes acts of exposing a sample to a plurality of primary nucleic acid probes (also called encoding probes); exposing the plurality of primary nucleic acid probes to a sequence of secondary nucleic acid probes (also called readout probes) and determining fluorescence of each of the secondary nucleic acid probes within the sample; creating codewords based on fluorescence of the secondary nucleic acid probes; and for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

In one set of embodiments, the method includes acts of exposing a plurality of primary nucleic acid probes to a sample; and exposing the plurality of nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary probes within the sample. In some embodiments, at least some of the plurality of secondary nucleic acid probes comprises distinguishable secondary nucleic acid probes

formed from combinatorial combination of one or more read sequences (or readout probe sequences) taken from a plurality of read sequences (or readout probe sequences).

In another set of embodiments, the method comprises acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a precision better than 500 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a precision better than 500 nm.

In another set of embodiments, the method comprises acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a resolution better than 100 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a resolution better than 100 nm.

The method, in yet another set of embodiments, includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity using a super-resolution imaging technique, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In certain embodiments, the method comprises acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords; and forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

In addition, in one set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes each comprising a target sequence; and forming groups comprising the

plurality of nucleic acid probes such that each group of nucleic acid probes corresponds to at least one common value of a position within the codewords.

In another set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions that is less than the number of targets, and wherein each codeword is associated with a single target, associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords, and forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

The method, in still another set of embodiments, includes acts of exposing a plurality of nucleic acid probes to a cell, exposing the plurality of nucleic acid probes to a sequence of secondary probes and determining fluorescence of each of the secondary probes within the cell, and based on the sequence of fluorescence of each of the secondary probes, determining nucleic acids within the cell.

In another set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes each comprising a target sequence; and forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes correspond to at least one common value of a position within the codewords.

In yet another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity using a super-resolution imaging technique, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In still another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a precision better than 500 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a precision better than 500 nm.

In still another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a resolution better than 100 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In still another set of embodiments, the method includes acts of associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code; associating unique read sequences with each possible value of each position in the codewords, wherein the read sequences are taken from a set of orthogonal sequences, which have limited homology with one another and with the nucleic acid species in a sample; forming a plurality of primary nucleic acid probes each comprising a target sequence that uniquely binds to a nucleic acid target and one or more read sequences; forming a plurality of secondary nucleic acid probes comprising a signaling entity and a sequence that is complementary to one of the read sequences; exposing a sample to the primary nucleic acid probes such that the nucleic acid probes hybridize to the nucleic acid targets in the sample; exposing the primary nucleic acid probes in the sample to a secondary nucleic acid probe such that the secondary nucleic acid probe hybridizes to the read sequence on at least some of the primary nucleic acid probes; imaging the sample; and repeating the exposing and imaging steps one or more times, using a different secondary nucleic acid probe for at least some of the repetitions.

The method, according to yet another set of embodiments, includes acts of associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes comprising a signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets; grouping the nucleic acid probes into a plurality of probe pools, wherein each of the probe pools corresponds to a specific value of a unique position within the codewords; exposing a sample to one of the probe pools; imaging the sample; and repeating the exposing and imaging steps one or more times, using a different probe pool for at least some of the repetitions.

In another aspect, the present invention encompasses methods of making one or more of the embodiments described herein. In still another aspect, the present invention encompasses methods of using one or more of the embodiments described herein.

Other advantages and novel features of the present invention will become apparent from the following detailed description of various non-limiting embodiments of the invention when considered in conjunction with the accompanying figures. In cases where the present specification and a document incorporated by reference include conflicting and/or inconsistent disclosure, the present specification shall control. If two or more documents incorporated by reference include conflicting and/or inconsistent disclosure with respect to each other, then the document having the later effective date shall control.

BRIEF DESCRIPTION OF THE DRAWINGS

Non-limiting embodiments of the present invention will be described by way of example with reference to the accompanying figures, which are schematic and are not intended to be drawn to scale. In the figures, each identical or nearly identical component illustrated is typically represented by a single numeral. For purposes of clarity, not every component is labeled in every figure, nor is every component of each embodiment of the invention shown where illustration is not necessary to allow those of ordinary skill in the art to understand the invention. In the figures:

Figs. 1A-1C illustrates an encoding scheme for nucleic acid probes, in certain embodiments of the invention;

Figs. 2A-2G illustrates determining of mRNAs in a cell, in some embodiments of the invention;

Figs. 3A-3B illustrate the determination of nucleic acids, in accordance with various embodiments of the invention;

Figs. 4A-4B is a non-limiting example of multiple read sequences distributed in a population of different nucleic acid probes, in accordance with certain embodiments of the invention;

Figs. 5A-5E illustrate the determination of nucleic acids, in accordance with another embodiment of the invention;

Figs. 6A-6H illustrate simultaneous determination of multiple nucleic acid species in cells, in certain embodiments of the invention;

Figs. 7A-7E show expression noise of genes and co-variation of expression between different genes determined in accordance with some embodiments of the invention;

Figs. 8A-8E illustrate spatial distribution of RNAs in cells determined in accordance with one embodiment of the invention;

Figs. 9A-9C illustrate simultaneous determination of multiple nucleic acid species in cells, in another embodiment of the invention;

Figs. 10A-10B show expression between different genes determined in accordance with yet another embodiment of the invention;

Fig. 11 is a schematic description of combinatorial labeling, in accordance with another embodiment of the invention;

Figs. 12A-12C show schematic descriptions of Hamming distance, in another embodiment of the invention;

Fig. 13 illustrates the production of a library of probes, in still another embodiment of the invention;

Figs. 14A-14B illustrate fluorescent spot determinations, in another embodiment of the invention;

Figs. 15A-15B illustrate error correction facilitates RNA detection, in yet another embodiment of the invention;

Figs. 16A-16E show characterization of misidentification rates and calling rates, in one embodiment of the invention;

Figs. 17A-17D show characterization of misidentification rates and calling rates, in another embodiment of the invention;

Figs. 18A-18C show a comparison of experiments, in accordance with one embodiment of the invention;

Figs. 19A-19D illustrate decoding and error assessment, in another embodiment of the invention; and

Figs. 20A to 20H illustrate encoding of 130 RNA species and 10 code words as misidentification controls as Codebook 1 and Codebook 2, in another embodiment of the invention.

BRIEF DESCRIPTION OF THE SEQUENCES

SEQ ID NO: 1 is:

GTTGGCGACGAAAGCACTGCGATTGGAACCGTCCCAAGCGTTGCGCTTAATGGA TCATCAATTTTGTCTCACTACGACGGTCAATCGCGCTGCATACTTGCGTCGGTCG GACAAACGAGG;

SEQ ID NO: 2 is CGCAACGCTTGGGACGGTTCCAATCGGATC;

SEQ ID NO: 3 is CGAATGCTCTGGCCTCGAACGAACGATAGC;

SEQ ID NO: 4 is ACAAATCCGACCAGATCGGACGATCATGGG;

SEQ ID NO: 5 is CAAGTATGCAGCGCGATTGACCGTCTCGTT;

SEQ ID NO: 6 is TGCGTCGTCTGGCTAGCACGGCACGCAAAT;

SEQ ID NO: 7 is AAGTCGTACGCCGATGCGCAGCAATTCACT;

SEQ ID NO: 8 is CGAAACATCGGCCACGGTCCCGTTGAACTT;

SEQ ID NO: 9 is ACGAATCCACCGTCCAGCGCGTCAAACAGA;

SEQ ID NO: 10 is CGCGAAATCCCCGTAACGAGCGTCCCTTGC;

SEQ ID NO: 11 is GCATGAGTTGCCTGGCGTTGCGACGACTAA;

SEQ ID NO: 12 is CCGTCGTCTCCGGTCCACCGTTGCGCTTAC;

SEQ ID NO: 13 is GGCCAATGGCCCAGGTCCGTCACGCAATTT;

SEQ ID NO: 14 is TTGATCGAATCGGAGCGTAGCGGAATCTGC;

SEQ ID NO: 15 is CGCGCGGATCCGCTTGTCGGGAACGGATAC;

SEQ ID NO: 16 is GCCTCGATTACGACGGATGTAATTCGGCCG;

SEQ ID NO: 17 is GCCCGTATTCCCGCTTGCGAGTAGGGCAAT

SEQ ID NO: 18 is GTTGGTCGGCACTTGGGTGC;

SEQ ID NO: 19 is CGATGCGCCAATTCCGGTTC;

SEQ ID NO: 20 is CGCGGGCTATATGCGAACCG;

SEQ ID NO: 21 is
TAATACGACTCACTATAGGGAAAGCCGGTTCATCCGGTGG;

SEQ ID NO: 22 is
TAATACGACTCACTATAGGGTGATCATCGCTCGCGGGTTG;

SEQ ID NO: 23 is
TAATACGACTCACTATAGGGCGTGGAGGGCATACAACGC;

SEQ ID NO: 24 is CGCAACGCTTGGGACGGTTCCAATCGGATC/3Cy5Sp/;

SEQ ID NO: 25 is CGAATGCTCTGGCCTCGAACGAACGATAGC/3Cy5Sp/;

SEQ ID NO: 26 is ACAAATCCGACCAGATCGGACGATCATGGG/3Cy5Sp/;

SEQ ID NO: 27 is CAAGTATGCAGCGCGATTGACCGTCTCGTT/3Cy5Sp/;

SEQ ID NO: 28 is GCGGGAAGCACGTGGATTAGGGCATCGACC/3Cy5Sp/;

SEQ ID NO: 29 is AAGTCGTACGCCGATGCGCAGCAATTCACT/3Cy5Sp/;

SEQ ID NO: 30 is CGAAACATCGGCCACGGTCCCGTTGAACTT/3Cy5Sp/;

SEQ ID NO: 31 is ACGAATCCACCGTCCAGCGCGTCAAACAGA/3Cy5Sp/;

SEQ ID NO: 32 is CGCGAAATCCCCGTAACGAGCGTCCCTTGC/3Cy5Sp/;

SEQ ID NO: 33 is GCATGAGTTGCCTGGCGTTGCGACGACTAA/3Cy5Sp/;

SEQ ID NO: 34 is CCGTCGTCTCCGGTCCACCGTTGCGCTTAC/3Cy5Sp/;

SEQ ID NO: 35 is GGCCAATGGCCCAGGTCCGTCACGCAATTT/3Cy5Sp/;
SEQ ID NO: 36 is TTGATCGAATCGGAGCGTAGCGGAATCTGC/3Cy5Sp/;
SEQ ID NO: 37 is CGCGCGGATCCGCTTGTCGGGAACGGATAC/3Cy5Sp/;
SEQ ID NO: 38 is GCCTCGATTACGACGGATGTAATTCGGCCG/3Cy5Sp/; and
SEQ ID NO: 39 is GCCCGTATTCCCGCTTGCGAGTAGGGCAAT/3Cy5Sp/

.

DETAILED DESCRIPTION

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. Certain embodiments are directed to determining nucleic acids, such as mRNA, within cells at relatively high resolutions. In some embodiments, a plurality of nucleic acid probes may be applied to a sample, and their binding within the sample determined, e.g., using fluorescence, to determine locations of the nucleic acid probes within the sample. In some embodiments, codewords may be based on the binding of the plurality of nucleic acid probes, and in some cases, the codewords may define an error-correcting code to reduce or prevent misidentification of the nucleic acids. In certain cases, a relatively large number of different targets may be identified using a relatively small number of labels, e.g., by using various combinatorial approaches.

Two example approaches are now discussed. It should be understood, however, that these are presented by way of explanation and not limitation; other aspects and embodiments are discussed in further detail herein. In one example method, primary probes (also called encoding probes) and secondary probes (also called readout probes) are used, where the primary probes encode "codewords" and bind to target nucleic acids in the sample, and the secondary probes are used to read out the codewords from the primary probes. In another example method, a plurality of different primary probes containing codewords are divided into as many separate pools as there are positions in the codewords, such that each primary probe pool corresponds to a certain value in a certain position of the codewords (e.g., a "one" in the first position as in "1001").

The first example is now described with respect to Fig. 3A. As will be discussed in more detail below, in other embodiments, other configurations may be used as well. In this first example, a series of nucleic acid probes are used to determine nucleic acids within a cell or other sample, e.g., qualitatively or quantitatively. For example, nucleic acids may be identified as being present or absent, and/or the numbers or concentrations of certain nucleic acids may be determined within the cell or other sample. In some cases, the positions of the

probes within the cell or other sample may be determined at relatively high resolutions, and in some cases, at resolutions better than the wavelength of visible light.

This example is generally directed to spatially detecting nucleic acids within a cell or other sample, e.g., at relatively high resolutions. For example, the nucleic acids may be mRNAs, or other nucleic acids described herein. In one set of embodiments, the nucleic acids within the cell may be determined by delivering or applying nucleic acid probes to the cell. In some cases, by using combinatorial approaches, a relatively large number of nucleic acids may be determined using a relatively small number of different labels on the nucleic acid probes. Thus, for example, a relatively small number of experiments may be used to determine a relatively large number of nucleic acids in a sample, e.g., due to simultaneous binding of the nucleic acid probes to different nucleic acids in the sample.

In one set of embodiments, a population of primary nucleic acid probes are applied to the cell (or other sample) that is able to bind nucleic acids suspected of being present within the cell. Afterwards, sequentially, secondary nucleic acid probes that can bind to or otherwise interact with some of the primary nucleic acids are added and determined, e.g., using imaging techniques such as fluorescence microscopy (e.g., conventional fluorescence microscopy), STORM (stochastic optical reconstruction microscopy) or other imaging techniques. After imaging, the secondary nucleic acid probes are inactivated or removed, and a different secondary nucleic acid probe is added to the sample. This may be repeated multiple times with multiple different secondary nucleic acid probes. The pattern of binding of the various secondary nucleic acid probes may be used to determine the primary nucleic acid probes at locations within the cell or other sample, which can be used to determine mRNA or other nucleic acids that are present.

For instance, as is shown in Fig. 3A, a population of nucleic acids 10 within a cell (represented here by nucleic acids 11, 12, and 13) may be exposed to a population of primary nucleic acid probes 20, including probes 21 and 22. The primary nucleic acid probes may contain, for instance, a target sequence that can recognize a nucleic acid (e.g., a sequence within nucleic acid 11). Probes 21 and 22 may contain the same or different targeting sequences, which may bind to or hybridize with the same or different nucleic acids. As an example, as is shown in Fig. 3A, probe 21 contains a first targeting sequence 25 which targets the probe to nucleic acid 11, while probe 22 contains a second targeting sequence 26, not identical to the first targeting sequence 25 and which targets the probe to nucleic acid 12. The target sequence may be substantially complementary to at least a portion of a target

nucleic acid, and enough of the target sequence may be present such that specific binding of the nucleic acid probe to the target nucleic acid can occur.

Primary nucleic acid probes 20 may also contain one or more "read" sequences. Two such read sequences are used in this example, although in other embodiments, there may be one, three, four, or more read sequences present within a primary nucleic acid probe. The read sequences may all independently be the same or different. In addition, in one set of embodiments, different nucleic acid probes may use one or more common read sequences. For example, more than one read sequence may be combinatorially combined on different nucleic acid probes, thereby producing a relatively large number of different nucleic acid probes that can be separately identified, even though only a relatively small number of read sequences are used. Thus, for example, in Fig. 3A, probe 21 contains read sequences 27 and 29, while probe 22 contains read sequences 27 and 28, where the two read sequences 27 are identical, and different from read sequences 28 and 29.

After primary nucleic acid probes 20 have been introduced to the sample and allowed to interact with nucleic acids 11, 12, and 13, one or more secondary nucleic acid probes 30 may be applied to the sample to determine the primary nucleic acid probes. The secondary nucleic acid probes may contain a recognition sequence able to recognize one of the read sequences present within the population of primary nucleic acid probes. For instance, the recognition sequence may be substantially complementary to at least a portion of the read sequence, such that the secondary nucleic acid probe is able to bind to or hybridize with corresponding primary nucleic acid probe. For instance, in this example, recognition sequence 35 is able to recognize read sequence 27. In addition, the secondary nucleic acid probes may contain one or more signaling entities 33. For example, a signaling entity may be a fluorescent entity attached to the probe, or a certain sequence of nucleic acids that can be determined in some fashion. More than one secondary sequence may be used, e.g., sequentially. For example, as shown in this figure, the initial secondary probe 30 may be removed (e.g., as discussed below) and a new secondary probe 31 may be added, containing recognition sequence 36 able to recognize read sequence 28 and one or more signaling entities 33. This may also be repeated multiple times, e.g., to determine read sequence 29 or other read sequences that may be present.

The location of the secondary nucleic acid probes 30, 31, etc. may be determined by determining signaling entity 33. For example, if the signaling entity is fluorescent, then fluorescence microscopy can be used to determine the signaling entity. In some embodiments, imaging of a sample to determine the signaling entity may be used at relatively

high resolutions, and in some cases, super-resolution imaging techniques (e.g., resolutions better than the wavelength of visible light or the diffraction limit of light) may be used. Examples of super-resolution imaging techniques include STORM, or other techniques as discussed herein. In some cases, e.g., with certain super-resolution imaging techniques such as STORM, more than one image of the sample may be acquired.

More than one type of secondary nucleic acid probe may be applied to a cell or other sample. For example, a first secondary nucleic acid probe may be applied that can recognize a first read sequence, then it or its attached signaling entity may be inactivated or removed, and a second secondary nucleic acid probe may be applied that can recognize a second read sequence. This process may be repeated multiple times, each with a different secondary nucleic acid probe, e.g., to determine the read sequences that were present in the various primary nucleic acid probes. Thus, primary nucleic acids within the sample can be determined on the basis of the binding pattern of secondary nucleic acid probes.

For example, a first location within the cell or other sample may exhibit binding of a first secondary probe and a third secondary probe, but not the binding of a second or a fourth secondary probe, while a second location may exhibit a different pattern of binding of various secondary probes. The primary nucleic acid probe that the secondary probes are able to bind to or hybridize with may be determined by considering the pattern of binding of various secondary probes. For instance, referring to Fig. 3A, if a first secondary probe is able to determine read sequence 27, a second secondary probe is able to determine read sequence 28, and a third secondary probe is able to determine read sequence 29, then primary nucleic acid 25 may be determined through the binding of the first and third secondary probes (but not the second secondary probe), while primary nucleic acid 26 may be determined through the binding of the first and second secondary probes (but not the third secondary probe). Similarly, if it is known that first probe 21 contains target sequence 25 while second probe 22 contains target sequence 26, then nucleic acids 11 and 12 may also be determined within the sample, e.g., spatially, based on the binding pattern of the various secondary nucleic acid probes. In addition, it should be noted that due to the presence of more than one read sequence on the primary nucleic acid probes, even though first probe 21 and second probe 22 contains a common read sequence (read sequence 27), these probes may be distinguished in the sample due to the different binding patterns of the various secondary nucleic acid probes.

In certain embodiments, this pattern of binding or hybridization of the secondary nucleic acid probes may be converted into a "codeword." In this example, for instance, the codewords are "101" and "110" for first probe 21 and second probe 22, respectively, where a

value of 1 represents binding and a value of 0 represents no binding. The codewords may also have longer lengths in other embodiments; only three probes are shown here for clarity purposes only. A codeword can be directly related to a specific target nucleic acid sequence of the primary nucleic acid probe. Accordingly, different primary nucleic acid probes may match certain codewords, which can then be used to identify the different targets of the primary nucleic acid probes based on the binding patterns of the secondary probes, even if in some cases, there is overlap in the read sequences of different secondary probes, e.g., as was shown in Fig. 3A. However, if no binding is evident (e.g., for nucleic acid 13), then the codeword would be "000" in this example.

The values in each codeword can also be assigned in different fashions in some embodiments. For example, a value of 0 could represent binding while a value of 1 represents no binding. Similarly, a value of 1 could represent binding of a secondary nucleic acid probe with one type of signaling entity while a value of 0 could represent binding of a secondary nucleic acid probe with another type of distinguishable signaling entity. These signaling entities could be distinguished, for example, via different colors of fluorescence. In some cases, values in codewords need not be confined to 0 and 1. The values could also be drawn from larger alphabets, such as ternary (e.g., 0, 1, and 2) or quaternary (e.g., 0, 1, 2, and 3) systems. Each different value could, for example, be represented by a different distinguishable signaling entity, including (in some cases) one value that may be represented by the absence of signal.

The codewords for each target may be assigned sequentially, or may be assigned at random. For instance, referring to Fig. 3A, a first nucleic acid target may be assigned to 101, while a second nucleic acid target may be assigned to 110. In addition, in some embodiments, the codewords may be assigned using an error-detection system or an error-correcting system, such as a Hamming system, a Golay code, or an extended Hamming system (or a SECDED system, i.e., single error correction, double error detection). Generally speaking, such systems can be used to identify where errors have occurred, and in some cases, such systems can also be used to correct the errors and determine what the correct codeword should have been. For example, a codeword such as 001 may be detected as invalid and corrected using such a system to 101, e.g., if 001 is not previously assigned to a different target sequence. A variety of different error-correcting codes can be used, many of which have previously been developed for use within the computer industry; however, such error-correcting systems have not typically been used within biological systems. Additional examples of such error-correcting codes are discussed in more detail below.

It should also be understood that all possible codewords in a code need not be used in some cases. For example, in some embodiments, codewords that are not used can serve as negative controls. Similarly, in some embodiments, some codewords can be left out because they are more prone to errors in measurement than other codewords. For example, in some implementations, reading a codeword with more values of '1' might be more error-prone that reading a codeword with fewer values of '1.'

It should be understood that the above description is an example of one embodiment of the invention, and that primary and secondary nucleic acid probes are not necessary in all embodiments. For example, in some embodiments, a series of nucleic acid probes containing signaling entities are used to determine nucleic acids within a cell or other sample, without necessarily requiring secondary probes.

For example, turning now to Fig. 3B, nucleic acids 11, 12, and 13 are exposed to different rounds of probes 21, 22, 23, 24, etc. in this example. These probes may each contain a target sequence that can recognize a nucleic acid (e.g., a sequence within nucleic acid 11 or 12). These probes may each target the same nucleic acid, but different regions of the nucleic acid. In addition, some or all of the probes may contain one or more signaling entities, e.g., signaling entity 29 on probe 21. For example, the signaling entity may be a fluorescent entity attached to the probe, or a certain sequence of nucleic acids that can be determined in some fashion.

The first round of probes (e.g. probe 21 and probe 22) may be applied to the cell or other sample. Probe 21 may be allowed to bind to nucleic acid 11 via target sequence 25. Such binding can be determined by determining signaling entity 29. For example, if the signaling entity is fluorescent, then fluorescence microscopy can be used to determine the signaling entity, e.g., spatially within the cell or other sample. In some but not all embodiments, imaging of a sample to determine the signaling entity may be used at relatively high resolutions, and in some cases, super-resolution imaging techniques may be used. Other, different probes may be present as well; for instance, probe 22 containing target sequence 26 may bind to nucleic acid 12, and be determined via signaling entity 29 within probe 22. These may occur, e.g., sequentially or simultaneously. Optionally, probes 21 and 22 may also be removed or inactivated, e.g., between application of different rounds of probes.

Next, a second round of probes (e.g., probe 23) is applied to the sample. In this example, probe 23 is able to bind to nucleic acid 11 via a targeting region, although there is no probe in the second round that is able to bind to nucleic acid 12. Binding of the probes is

allowed to occur as discussed above, and determination of binding may occur via signaling entities. These signaling entities may be the same or different as from the first round of probes. This process may be repeated any number of times with different probes. For example, as is shown in Fig. 3B, round 2 contains probes able to bind to nucleic acid 11, while round 3 contains probes able to bind to nucleic acid 12.

In certain embodiments, each round of binding or hybridization of nucleic acid probes may be converted into a "codeword." In this example, using probes 21, 22, 23, and 24, the codewords 101 or 110 could be formed, where 1 represents binding and 0 represents no binding and the first position corresponds to the binding of probes 21 or 22 while the second position corresponds to the binding of probes 22, and the third position corresponds to the binding of probe 24. A codeword of 000 would represent no binding, e.g., as shown with nucleic acid 13 in this example. A codeword can be directly related to a specific target nucleic acid sequence of the nucleic acid probes, by designing appropriate nucleic acid probes. Thus, for example, 110 may correspond to a first target nucleic acid 12 (e.g., the first and second round of nucleic acid probes containing probes able to target nucleic acid 11, and these probes may target the same or different regions of nucleic acid 11) while 101 may correspond to a second target nucleic acid (e.g., the first and third round of nucleic acid probes containing probes able to target nucleic acid 12, and these probes may target the same or different regions of nucleic acid 12). In addition, it should be noted that each round of probes may contain the same, or different signaling entities as other probes in the same round, and/or other probes in different rounds. For instance, in one set of embodiments, only one signaling entity is used in all of the rounds of probes.

Similar to the above, the codewords for each target may be assigned sequentially, or may be assigned at random. The codewords may be assigned within a code space in some embodiments using an error-detection or an error-correcting system, such as a Hamming system, a Golay code, or an extended Hamming system or a SECDED system (single error correction, double error detection). Generally speaking, such error-correction systems can be used to identify where errors have occurred, and in some cases, such systems can also be used to correct the errors and determine what the correct codeword should have been.

Similar to the above, the values at each position in the codeword can be arbitrarily assigned in certain embodiments to binding or non-binding of probes that contain more than one distinguishable signaling entity.

In some cases, the nucleic acid probes may be formed into "pools" or groups of nucleic acids that share a common feature. For example, probes to all targets with codewords

that contain a 1 in the first position, e.g. 110 and 101 but not 011, may comprise one pool while probes to all targets that contain a 1 in the second position, e.g. 110 and 011 but not 101, may comprise another pool. See also Fig. 1C. In some cases, a nucleic acid probe may be a member of more than one group or pool. Members of a nucleic acid pool may also contain features in addition to target sequences, read sequences, and or signaling entities that allow them to be distinguished from other groups. These features may be short nucleic acid sequences that are used for the amplification, production, or separation of these sequences. The nucleic acid probes of each group may be applied to a sample, e.g., sequentially, as discussed herein.

Thus, in some aspects, the present invention is generally directed to systems and methods for determining nucleic acids within a cell or other sample. The sample may include a cell culture, a suspension of cells, a biological tissue, a biopsy, an organism, or the like. The sample may also be cell-free but nevertheless contain nucleic acids. If the sample contains a cell, the cell may be a human cell, or any other suitable cell, e.g., a mammalian cell, a fish cell, an insect cell, a plant cell, or the like. More than one cell may be present in some cases.

The nucleic acids to be determined may be, for example, DNA, RNA, or other nucleic acids that are present within a cell (or other sample). The nucleic acids may be endogenous to the cell, or added to the cell. For instance, the nucleic acid may be viral, or artificially created. In some cases, the nucleic acid to be determined may be expressed by the cell. The nucleic acid is RNA in some embodiments. The RNA may be coding and/or non-coding RNA. Non-limiting examples of RNA that may be studied within the cell include mRNA, siRNA, rRNA, miRNA, tRNA, lncRNA, snoRNAs, snRNAs, exRNAs, piRNAs, or the like.

In some cases, a significant portion of the nucleic acid within the cell may be studied. For instance, in some cases, enough of the RNA present within a cell may be determined so as to produce a partial or complete transcriptome of the cell. In some cases, at least 4 types of mRNAs are determined within a cell, and in some cases, at least 3, at least 4, at least 7, at least 8, at least 12, at least 14, at least 15, at least 16, at least 22, at least 30, at least 31, at least 32, at least 50, at least 63, at least 64, at least 72, at least 75, at least 100, at least 127, at least 128, at least 140, at least 255, at least 256, at least 500, at least 1,000, at least 1,500, at least 2,000, at least 2,500, at least 3,000, at least 4,000, at least 5,000, at least 7,500, at least 10,000, at least 12,000, at least 15,000, at least 20,000, at least 25,000, at least 30,000, at least 40,000, at least 50,000, at least 75,000, or at least 100,000 types of mRNAs may be determined within a cell.

In some cases, the transcriptome of a cell may be determined. It should be understood that the transriptome generally encompasses all RNA molecules produced within a cell, not just mRNA. Thus, for instance, the transcriptome may also include rRNA, tRNA, siRNA, etc. In some embodiments, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or 100% of the transcriptome of a cell may be determined.

The determination of one or more nucleic acids within the cell or other sample may be qualitative and/or quantitative. In addition, the determination may also be spatial, e.g., the position of the nucleic acid within the cell or other sample may be determined in two or three dimensions. In some embodiments, the positions, number, and/or concentrations of nucleic acids within the cell (or other sample) may be determined.

In some cases, a significant portion of the genome of a cell may be determined. The determined genomic segments may be continuous or interspersed on the genome. For example, in some cases, at least 4 genomic segments are determined within a cell, and in some cases, at least 3, at least 4, at least 7, at least 8, at least 12, at least 14, at least 15, at least 16, at least 22, at least 30, at least 31, at least 32, at least 50, at least 63, at least 64, at least 72, at least 75, at least 100, at least 127, at least 128, at least 140, at least 255, at least 256, at least 500, at least 1,000, at least 1,500, at least 2,000, at least 2,500, at least 3,000, at least 4,000, at least 5,000, at least 7,500, at least 10,000, at least 12,000, at least 15,000, at least 20,000, at least 25,000, at least 30,000, at least 40,000, at least 50,000, at least 75,000, or at least 100,000 genomic segments may be determined within a cell.

In some cases, the entire genome of a cell may be determined. It should be understood that the genome generally encompasses all DNA molecules produced within a cell, not just chromosome DNA. Thus, for instance, the genome may also include, in some cases, mitochondria DNA, chloroplast DNA, plasmid DNA, etc. In some embodiments, at least about 5%, at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or 100% of the genome of a cell may be determined.

As discussed herein, a variety of nucleic acid probes may be used to determine one or more nucleic acids within a cell or other sample. The probes may comprise nucleic acids (or entities that can hybridize to a nucleic acid, e.g., specifically) such as DNA, RNA, LNA (locked nucleic acids), PNA (peptide nucleic acids), or combinations thereof. In some cases,

additional components may also be present within the nucleic acid probes, e.g., as discussed below. Any suitable method may be used to introduce nucleic acid probes into a cell.

For example, in some embodiments, the cell is fixed prior to introducing the nucleic acid probes, e.g., to preserve the positions of the nucleic acids within the cell. Techniques for fixing cells are known to those of ordinary skill in the art. As non-limiting examples, a cell may be fixed using chemicals such as formaldehyde, paraformaldehyde, glutaraldehyde, ethanol, methanol, acetone, acetic acid, or the like. In one embodiment, a cell may be fixed using Hepes-glutamic acid buffer-mediated organic solvent (HOPE).

The nucleic acid probes may be introduced into the cell (or other sample) using any suitable method. In some cases, the cell may be sufficiently permeabilized such that the nucleic acid probes may be introduced into the cell by flowing a fluid containing the nucleic acid probes around the cells. In some cases, the cells may be sufficiently permeabilized as part of a fixation process; in other embodiments, cells may be permeabilized by exposure to certain chemicals such as ethanol, methanol, Triton, or the like. In addition, in some embodiments, techniques such as electroporation or microinjection may be used to introduce nucleic acid probes into a cell or other sample.

Certain aspects of the present invention are generally directed to nucleic acid probes that are introduced into a cell (or other sample). The probes may comprise any of a variety of entities that can hybridize to a nucleic acid, typically by Watson-Crick base pairing, such as DNA, RNA, LNA, PNA, etc., depending on the application. The nucleic acid probe typically contains a target sequence that is able to bind to at least a portion of a target nucleic acid, in some cases specifically. When introduced into a cell or other system, the target system may be able to bind to a specific target nucleic acid (e.g., an mRNA, or other nucleic acids as discussed herein). In some cases, the nucleic acid probes may be determined using signaling entities (e.g., as discussed below), and/or by using secondary nucleic acid probes able to bind to the nucleic acid probes (i.e., to primary nucleic acid probes). The determination of such nucleic acid probes is discussed in detail below.

In some cases, more than one type of (primary) nucleic acid probe may be applied to a sample, e.g., simultaneously. For example, there may be at least 2, at least 5, at least 10, at least 25, at least 50, at least 75, at least 100, at least 300, at least 1,000, at least 3,000, at least 10,000, or at least 30,000 distinguishable nucleic acid probes that are applied to a sample, e.g., simultaneously or sequentially.

The target sequence may be positioned anywhere within the nucleic acid probe (or primary nucleic acid probe or encoding nucleic acid probe). The target sequence may contain

a region that is substantially complementary to a portion of a target nucleic acid. In some cases, the portions may be at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% complementary. In some cases, the target sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 50, at least 60, at least 65, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides in length. In some cases, the target sequence may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the target sequence may have a length of between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc. Typically, complementarity is determined on the basis of Watson-Crick nucleotide base pairing.

The target sequence of a (primary) nucleic acid probe may be determined with reference to a target nucleic acid suspected of being present within a cell or other sample. For example, a target nucleic acid to a protein may be determined using the protein's sequence, by determining the nucleic acids that are expressed to form the protein. In some cases, only a portion of the nucleic acids encoding the protein are used, e.g., having the lengths as discussed above. In addition, in some cases, more than one target sequence that can be used to identify a particular target may be used. For instance, multiple probes can be used, sequentially and/or simultaneously, that can bind to or hybridize to different regions of the same target. Hybridization typically refers to an annealing process by which complementary single-stranded nucleic acids associate through Watson-Crick nucleotide base pairing (e.g., hydrogen bonding, guanine-cytosine and adenine-thymine) to form double-stranded nucleic acid.

In some embodiments, a nucleic acid probe, such as a primary nucleic acid probe, may also comprise one or more "read" sequences. However, it should be understood that read sequences are not necessary in all cases. In some embodiments, the nucleic acid probe may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 or more, 20 or more, 32 or more, 40 or more, 50 or more, 64 or more, 75 or more, 100 or more, 128 or more read

sequences. The read sequences may be positioned anywhere within the nucleic acid probe. If more than one read sequence is present, the read sequences may be positioned next to each other, and/or interspersed with other sequences.

The read sequences, if present, may be of any length. If more than one read sequence is used, the read sequences may independently have the same or different lengths. For instance, the read sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 50, at least 60, at least 65, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides in length. In some cases, the read sequence may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the read sequence may have a length of between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc.

The read sequence may be arbitrary or random in some embodiments. In certain cases, the read sequences are chosen so as to reduce or minimize homology with other components of the cell or other sample, e.g., such that the read sequences do not themselves bind to or hybridize with other nucleic acids suspected of being within the cell or other sample. In some cases, the homology may be less than 10%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, or less than 1%. In some cases, there may be a homology of less than 20 basepairs, less than 18 basepairs, less than 15 basepairs, less than 14 basepairs, less than 13 basepairs, less than 12 basepairs, less than 11 basepairs, or less than 10 basepairs. In some cases, the basepairs are sequential.

In one set of embodiments, a population of nucleic acid probes may contain a certain number of read sequences, which may be less than the number of targets of the nucleic acid probes in some cases. Those of ordinary skill in the art will be aware that if there is one signaling entity and n read sequences, then in general 2^n-1 different nucleic acid targets may be uniquely identified. However, not all possible combinations need be used. For instance, a population of nucleic acid probes may target 12 different nucleic acid sequences, yet contain no more than 8 read sequences. As another example, a population of nucleic acids may target 140 different nucleic acid species, yet contain no more than 16 read sequences. Different

nucleic acid sequence targets may be separately identified by using different combinations of read sequences within each probe. For instance, each probe may contain 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, etc. or more read sequences. In some cases, a population of nucleic acid probes may each contain the same number of read sequences, although in other cases, there may be different numbers of read sequences present on the various probes.

As a non-limiting example, a first nucleic acid probe may contain a first target sequence, a first read sequence, and a second read sequence, while a second, different nucleic acid probe may contain a second target sequence, the same first read sequence, but a third read sequence instead of the second read sequence. Such probes may thereby be distinguished by determining the various read sequences present or associated with a given probe or location, as discussed herein.

In addition, the nucleic acid probes (and their corresponding, complimentary sites on the encoding probes), in certain embodiments, may be made using only 2 or only 3 of the 4 bases, such as leaving out all the "G"s or leaving out all of the "C"s within the probe. Sequences lacking either "G"s or "C"s may form very little secondary structure in certain embodiments, and can contribute to more uniform, faster hybridization.

In some embodiments, the nucleic acid probe may contain a signaling entity. It should be understood that signaling entities are not required in all cases, however; for instance, the nucleic acid probe may be determined using secondary nucleic acid probes in some embodiments, as is discussed in additional detail below. Examples of signaling entities that can be used are also discussed in more detail below.

Other components may also be present within a nucleic acid probe as well. For example, in one set of embodiments, one or more primer sequences may be present, e.g., to allow for enzymatic amplification of probes. Those of ordinary skill in the art will be aware of primer sequences suitable for applications such as amplification (e.g., using PCR or other suitable techniques). Many such primer sequences are available commercially. Other examples of sequences that may be present within a primary nucleic acid probe include, but are not limited to promoter sequences, operons, identification sequences, nonsense sequences, or the like.

Typically, a primer is a single-stranded or partially double-stranded nucleic acid (e.g., DNA) that serves as a starting point for nucleic acid synthesis, allowing polymerase enzymes such as nucleic acid polymerase to extend the primer and replicate the complementary strand. A primer is (e.g., is designed to be) complementary to and to hybridize to a target nucleic acid. In some embodiments, a primer is a synthetic primer. In some embodiments, a primer

is a non-naturally-occurring primer. A primer typically has a length of 10 to 50 nucleotides. For example, a primer may have a length of 10 to 40, 10 to 30, 10 to 20, 25 to 50, 15 to 40, 15 to 30, 20 to 50, 20 to 40, or 20 to 30 nucleotides. In some embodiments, a primer has a length of 18 to 24 nucleotides.

In addition, the components of the nucleic acid probe may be arranged in any suitable order. For instance, in one embodiment, the components may be arranged in a nucleic acid probe as: primer—read sequences—targeting sequence—read sequences—reverse primer. The "read sequences" in this structure may each contain any number (including 0) of read sequences, so long as at least one read sequence is present in the probe. Non-limiting example structures include primer—targeting sequence—read sequences—reverse primer, primer—read sequences—targeting sequence—reverse primer, targeting sequence—primer—targeting sequence—read sequences—reverse primer, targeting sequence—primer—read sequences—targeting sequence—reverse primer, primer—target sequence—read sequences—targeting sequence—reverse primer, targeting sequence—primer—read sequence—reverse primer, targeting sequence—read sequence—primer, read sequence—targeting sequence—primer, read sequence—primer—targeting sequence—reverse primer, etc. In addition, the reverse primer is optional in some embodiments, including in all of the above-described examples.

After introduction of the nucleic acid probes into a cell or other sample, the nucleic acid probes may be directly determined by determining signaling entities (if present), and/or the nucleic acid probes may be determined by using one or more secondary nucleic acid probes, in accordance with certain aspects of the invention. As mentioned, in some cases, the determination may be spatial, e.g., in two or three dimensions. In addition, in some cases, the determination may be quantitative, e.g., the amount or concentration of a primary nucleic acid probe (and of a target nucleic acid) may be determined. Additionally, the secondary probes may comprise any of a variety of entities able to hybridize a nucleic acid, e.g., DNA, RNA, LNA, and/or PNA, etc., depending on the application. Signaling entities are discussed in more detail below.

A secondary nucleic acid probe may contain a recognition sequence able to bind to or hybridize with a read sequence of a primary nucleic acid probe. In some cases, the binding is specific, or the binding may be such that a recognition sequence preferentially binds to or hybridizes with only one of the read sequences that are present. The secondary nucleic acid probe may also contain one or more signaling entities. If more than one secondary nucleic acid probe is used, the signaling entities may be the same or different.

The recognition sequences may be of any length, and multiple recognition sequences may be of the same or different lengths. If more than one recognition sequence is used, the recognition sequences may independently have the same or different lengths. For instance, the recognition sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, or at least 50 nucleotides in length. In some cases, the recognition sequence may be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the recognition sequence may have a length of between 10 and 30, between 20 and 40, or between 25 and 35 nucleotides, etc. In one embodiment, the recognition sequence is of the same length as the read sequence. In addition, in some cases, the recognition sequence may be at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 100% complementary to a read sequence of the primary nucleic acid probe.

As mentioned, in some cases, the secondary nucleic acid probe may comprise one or more signaling entities. Examples of signaling entities are discussed in more detail below.

As discussed, in certain aspects of the invention, nucleic acid probes are used that contain various "read sequences." For example, a population of primary nucleic acid probes may contain certain "read sequences" which can bind certain of the secondary nucleic acid probes, and the locations of the primary nucleic acid probes are determined within the sample using secondary nucleic acid probes, e.g., which comprise a signaling entity. As mentioned, in some cases, a population of read sequences may be combined in various combinations to produce different nucleic acid probes, e.g., such that a relatively small number of read sequences may be used to produce a relatively large number of different nucleic acid probes.

Thus, in some cases, a population of primary nucleic acid probes (or other nucleic acid probes) may each contain a certain number of read sequences, some of which are shared between different primary nucleic acid probes such that the total population of primary nucleic acid probes may contain a certain number of read sequences. A population of nucleic acid probes may have any suitable number of read sequences. For example, a population of primary nucleic acid probes may have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 etc. read sequences. More than 20 are also possible in some embodiments. In addition, in some cases, a population of nucleic acid probes may, in total, have 1 or more, 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or

more, 11 or more, 12 or more, 13 or more, 14 or more, 15 or more, 16 or more, 20 or more, 24 or more, 32 or more, 40 or more, 50 or more, 60 or more, 64 or more, 100 or more, 128 or more, etc. of possible read sequences present, although some or all of the probes may each contain more than one read sequence, as discussed herein. In addition, in some embodiments, the population of nucleic acid probes may have no more than 100, no more than 80, no more than 64, no more than 60, no more than 50, no more than 40, no more than 32, no more than 24, no more than 20, no more than 16, no more than 15, no more than 14, no more than 13, no more than 12, no more than 11, no more than 10, no more than 9, no more than 8, no more than 7, no more than 6, no more than 5, no more than 4, no more than 3, or no more than two read sequences present. Combinations of any of these are also possible, e.g., a population of nucleic acid probes may comprise between 10 and 15 read sequences in total.

As a non-limiting example of an approach to combinatorially producing a relatively large number of nucleic acid probes from a relatively small number of read sequences, in a population of 6 different types of nucleic acid probes, each comprising one or more read sequences, the total number of read sequences within the population may be no greater than 4. It should be understood that although 4 read sequences are used in this example for ease of explanation, in other embodiments, larger numbers of nucleic acid probes may be realized, for example, using 5, 8, 10, 16, 32, etc. or more read sequences, or any other suitable number of read sequences described herein, depending on the application. Referring now to Fig. 4A, if each of the primary nucleic acid probes contains two different read sequences, then by using 4 such read sequences (A, B, C, and D), up to 6 probes may be separately identified. It should be noted that in this example, the ordering of read sequences on a nucleic acid probe is not essential, i.e., "AB" and "BA" may be treated as being synonymous (although in other embodiments, the ordering of read sequences may be essential and "AB" and "BA" may not necessarily be synonymous). Similarly, if 5 read sequences are used (A, B, C, D, and E) in the population of primary nucleic acid probes, up to 10 probes may be separately identified, as is shown in Fig. 4B. For example, one of ordinary skill in the art would understand that, for k read sequences in a population with n read sequences on each probe, up to $\binom{n}{k}$ different probes may be produced, assuming that the ordering of read sequences is not essential; because not all of the probes need to have the same number of read sequences and not all combinations of read sequences need to be used in every embodiment, either more or less than this number of different probes may also be used in certain embodiments. In addition, it should also be understood that the number of read sequences on each probe need not be

identical in some embodiments. For instance example, some probes may contain 2 read sequences while other probes may contain 3 read sequences.

In some aspects, the read sequences and/or the pattern of binding of nucleic acid probes within a sample may be used to define an error-detecting and/or an error-correcting code, for example, to reduce or prevent misidentification or errors of the nucleic acids, e.g., as was discussed with reference to Fig. 3. Thus, for example, if binding is indicated (e.g., as determined using a signaling entity), then the location may be identified with a "1"; conversely, if no binding is indicated, then the location may be identified with a "0" (or vice versa, in some cases). Multiple rounds of binding determinations, e.g., using different nucleic acid probes, can then be used to create a "codeword," e.g., for that spatial location. In some embodiments, the codeword may be subjected to error detection and/or correction. For instance, the codewords may be organized such that, if no match is found for a given set of read sequences or binding pattern of nucleic acid probes, then the match may be identified as an error, and optionally, error correction may be applied sequences to determine the correct target for the nucleic acid probes. In some cases, the codewords may have fewer "letters" or positions that the total number of nucleic acids encoded by the codewords, e.g. where each codeword encodes a different nucleic acid.

Such error-detecting and/or the error-correction code may take a variety of forms. A variety of such codes have previously been developed in other contexts such as the telecommunications industry, such as Golay codes or Hamming codes. In one set of embodiments, the read sequences or binding patterns of the nucleic acid probes are assigned such that not every possible combination is assigned.

For example, if 4 read sequences are possible and a primary nucleic acid probe contains 2 read sequences, then up to 6 primary nucleic acid probes could be identified; but the number of primary nucleic acid probes used may be less than 6. Similarly, for k read sequences in a population with n read sequences on each primary nucleic acid probe, $\binom{n}{k}$ different probes may be produced, but the number of primary nucleic acid probes that are used may be any number more or less than $\binom{n}{k}$. In addition, these may be randomly assigned, or assigned in specific ways to increase the ability to detect and/or correct errors.

As another example, if multiple rounds of nucleic acid probes are used, the number of rounds may be arbitrarily chosen. If in each round, each target can give two possible outcomes, such as being detected or not being detected, up to 2^n different targets may be possible for n rounds of probes, but the number of nucleic acid targets that are actually used

may be any number less than 2^n. For example, if in each round, each target can give more than two possible outcomes, such as being detected in different color channels, more than 2^n (e.g. 3^n, 4^n ...) different targets may be possible for n rounds of probes. In some cases, the number of nucleic acid targets that are actually used may be any number less than this number. In addition, these may be randomly assigned, or assigned in specific ways to increase the ability to detect and/or correct errors.

For example, in one set of embodiments, the codewords or nucleic acid probes may be assigned within a code space such that the assignments are separated by a Hamming distance, which measures the number of incorrect "reads" in a given pattern that cause the nucleic acid probe to be misinterpreted as a different valid nucleic acid probe. In certain cases, the Hamming distance may be at least 2, at least 3, at least 4, at least 5, at least 6, or the like. In addition, in one set of embodiments, the assignments may be formed as a Hamming code, for instance, a Hamming(7, 4) code, a Hamming(15, 11) code, a Hamming(31, 26) code, a Hamming(63, 57) code, a Hamming(127, 120) code, etc. In another set of embodiments, the assignments may form a SECDED code, e.g., a SECDED(8,4) code, a SECDED(16,4) code, a SCEDED(16, 11) code, a SCEDED(22, 16) code, a SCEDED(39, 32) code, a SCEDED(72, 64) code, etc. In yet another set of embodiments, the assignments may form an extended binary Golay code, a perfect binary Golay code, or a ternary Golay code. In another set of embodiments, the assignments may represent a subset of the possible values taken from any of the codes described above.

For example, a code with the same error correcting properties of the SECDED code may be formed by using only binary words that contain a fixed number of '1' bits, such as 4, to encode the targets. In another set of embodiments, the assignments may represent a subset of the possible values taken from codes described above for the purpose of addressing asymmetric readout errors. For example, in some cases, a code in which the number of '1' bits may be fixed for all used binary words may eliminate the biased measurement of words with different numbers of '1's when the rate at which '0' bits are measured as '1's or '1' bits are measured as '0's are different.

Accordingly, in some embodiments, once the codeword is determined (e.g., as discussed herein), the codeword may be compared to the known nucleic acid codewords. If a match is found, then the nucleic acid target can be identified or determined. If no match is found, then an error in the reading of the codeword may be identified. In some cases, error correction can also be applied to determine the correct codeword, and thus resulting in the correct identity of the nucleic acid target. In some cases, the codewords may be selected such

that, assuming that there is only one error present, only one possible correct codeword is available, and thus, only one correct identity of the nucleic acid target is possible. In some cases, this may also be generalized to larger codeword spacings or Hamming distances; for instance, the codewords may be selected such that if two, three, or four errors are present (or more in some cases), only one possible correct codeword is available, and thus, only one correct identity of the nucleic acid targets is possible.

The error-correcting code may be a binary error-correcting code, or it may be based on other numbering systems, e.g., ternary or quaternary error-correcting codes. For instance, in one set of embodiments, more than one type of signaling entity may be used and assigned to different numbers within the error-correcting code. Thus, as a non-limiting example, a first signaling entity (or more than one signaling entity, in some cases) may be assigned as "1" and a second signaling entity (or more than one signaling entity, in some cases) may be assigned as "2" (with "0" indicating no signaling entity present), and the codewords distributed to define a ternary error-correcting code. Similarly, a third signaling entity may additionally be assigned as "3" to make a quaternary error-correcting code, etc.

As discussed above, in certain aspects, signaling entities are determined, e.g., to determine nucleic acid probes and/or to create codewords. In some cases, signaling entities within a sample may be determined, e.g., spatially, using a variety of techniques. In some embodiments, the signaling entities may be fluorescent, and techniques for determining fluorescence within a sample, such as fluorescence microscopy or confocal microscopy, may be used to spatially identify the positions of signaling entities within a cell. In some cases, the positions of entities within the sample may be determined in two or even three dimensions. In addition, in some embodiments, more than one signaling entity may be determined at a time (e.g., signaling entities with different colors or emissions), and/or sequentially.

In addition, in some embodiments, a confidence level for the identified nucleic acid target may be determined. For example, the confidence level may be determined using a ratio of the number of exact matches to the number of matches having one or more one-bit errors. In some cases, only matches having a confidence ratio greater than a certain value may be used. For instance, in certain embodiments, matches may be accepted only if the confidence ratio for the match is greater than about 0.01, greater than about 0.03, greater than about 0.05, greater than about 0.1, greater than about 0.3, greater than about 0.5, greater than about 1, greater than about 3, greater than about 5, greater than about 10, greater than about 30, greater than about 50, greater than about 100, greater than about 300, greater than about

500, greater than about 1000, or any other suitable value. In addition, in some embodiments, matches may be accepted only if the confidence ratio for the identified nucleic acid target is greater than an internal standard or false positive control by about 0.01, about 0.03, about 0.05, about 0.1, about 0.3, about 0.5, about 1, about 3, about 5, about 10, about 30, about 50, about 100, about 300, about 500, about 1000, or any other suitable value

In some embodiments, the spatial positions of the entities (and thus, nucleic acid probes that the entities may be associated with) may be determined at relatively high resolutions. For instance, the positions may be determined at spatial resolutions of better than about 100 micrometers, better than about 30 micrometers, better than about 10 micrometers, better than about 3 micrometers, better than about 1 micrometer, better than about 800 nm, better than about 600 nm, better than about 500 nm, better than about 400 nm, better than about 300 nm, better than about 200 nm, better than about 100 nm, better than about 90 nm, better than about 80 nm, better than about 70 nm, better than about 60 nm, better than about 50 nm, better than about 40 nm, better than about 30 nm, better than about 20 nm, or better than about 10 nm, etc.

There are a variety of techniques able to determine or image the spatial positions of entities optically, e.g., using fluorescence microscopy. In some cases, the spatial positions may be determined at super resolutions, or at resolutions better than the wavelength of light or the diffraction limit. Non-limiting examples include STORM (stochastic optical reconstruction microscopy), STED (stimulated emission depletion microscopy), NSOM (Near-field Scanning Optical Microscopy), 4Pi microscopy, SIM (Structured Illumination Microscopy), SMI (Spatially Modulated Illumination) microscopy, RESOLFT (Reversible Saturable Optically Linear Fluorescence Transition Microscopy), GSD (Ground State Depletion Microscopy), SSIM (Saturated Structured-Illumination Microscopy), SPDM (Spectral Precision Distance Microscopy), Photo-Activated Localization Microscopy (PALM), Fluorescence Photoactivation Localization Microscopy (FPALM), LIMON (3D Light Microscopical Nanosizing Microscopy), Super-resolution optical fluctuation imaging (SOFI), or the like. See, e.g., U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*; or Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*, each incorporated herein by reference in their entireties.

As an illustrative non-limiting example, in one set of embodiments, the sample may be imaged with a high numerical aperture, oil immersion objective with 100X magnification and light collected on an electron-multiplying CCD camera. In another example, the sample could be imaged with a high numerical aperture, oil immersion lens with 40X magnification and light collected with a wide-field scientific CMOS camera. With different combinations of objectives and cameras, a single field of view may correspond to no less than 40 x 40 microns, 80 x 80 microns, 120 x 120 microns, 240 x 240 microns, 340 x 340 microns, or 500 x 500 microns, etc. in various non-limiting embodiments. Similarly, a single camera pixel may correspond, in some embodiments, to regions of the sample of no less than 80x80 nm, 120x120 nm, 160x160 nm, 240x240 nm, or 300x300 nm, etc. In another example, the sample may be imaged with a low numerical aperture, air lens with 10X magnification and light collected with a sCMOS camera. In additional embodiments, the sample may be optically sectioned by illuminating it via a single or multiple scanned diffraction limited foci generated either by scanning mirrors or a spinning disk and the collected passed through a single or multiple pinholes. In another embodiment, the sample may also be illuminated via thin sheet of light generated via any one of multiple methods known to those versed in the art.

In one embodiment, the sample may be illuminated by single Gaussian mode laser lines. In some embodiments, the illumination profiled may be flattened by passing these laser lines through a multimode fiber that is vibrated via piezo-electric or other mechanical means. In some embodiments, the illumination profile may be flattened by passing single-mode, Gaussian beams through a variety of refractive beam shapers, such as the piShaper or a series of stacked Powell lenses. In yet another set of embodiments, the Gaussian beams may be passed through a variety of different diffusing elements, such as ground glass or engineered diffusers, which may be spun in some cases at high speeds to remove residual laser speckle. In yet another embodiment, laser illumination may be passed through a series of lenslet arrays to produce overlapping images of the illumination that approximate a flat illumination field.

In some embodiments, the centroids of the spatial positions of the entities may be determined. For example, a centroid of a signaling entity may be determined within an image or series of images using image analysis algorithms known to those of ordinary skill in the art. In some cases, the algorithms may be selected to determine non-overlapping single emitters and/or partially overlapping single emitters in a sample. Non-limiting examples of suitable techniques include a maximum likelihood algorithm, a least squares algorithm, a

Bayesian algorithm, a compressed sensing algorithm, or the like. Combinations of these techniques may also be used in some cases.

In addition, the signaling entity may be inactivated in some cases. For example, in some embodiments, a first secondary nucleic acid probe containing a signaling entity may be applied to a sample that can recognize a first read sequence, then the first secondary nucleic acid probe can be inactivated before a second secondary nucleic acid probe is applied to the sample. If multiple signaling entities are used, the same or different techniques may be used to inactivate the signaling entities, and some or all of the multiple signaling entities may be inactivated, e.g., sequentially or simultaneously.

Inactivation may be caused by removal of the signaling entity (e.g., from the sample, or from the nucleic acid probe, etc.), and/or by chemically altering the signaling entity in some fashion, e.g., by photobleaching the signaling entity, bleaching or chemically altering the structure of the signaling entity, e.g., by reduction, etc.). For instance, in one set of embodiments, a fluorescent signaling entity may be inactivated by chemical or optical techniques such as oxidation, photobleaching, chemically bleaching, stringent washing or enzymatic digestion or reaction by exposure to an enzyme, dissociating the signaling entity from other components (e.g., a probe), chemical reaction of the signaling entity (e.g., to a reactant able to alter the structure of the signaling entity) or the like. For instance, bleaching may occur by exposure to oxygen, reducing agents, or the signaling entity could be chemically cleaved from the nucleic acid probe and washed away via fluid flow.

In some embodiments, various nucleic acid probes (including primary and/or secondary nucleic acid probes) may include one or more signaling entities. If more than one nucleic acid probe is used, the signaling entities may each by the same or different. In certain embodiments, a signaling entity is any entity able to emit light. For instance, in one embodiment, the signaling entity is fluorescent. In other embodiments, the signaling entity may be phosphorescent, radioactive, absorptive, etc. In some cases, the signaling entity is any entity that can be determined within a sample at relatively high resolutions, e.g., at resolutions better than the wavelength of visible light or the diffraction limit. The signaling entity may be, for example, a dye, a small molecule, a peptide or protein, or the like. The signaling entity may be a single molecule in some cases. If multiple secondary nucleic acid probes are used, the nucleic acid probes may comprise the same or different signaling entities.

Non-limiting examples of signaling entities include fluorescent entities (fluorophores) or phosphorescent entities, for example, cyanine dyes (e.g., Cy2, Cy3, Cy3B, Cy5, Cy5.5,

Cy7, etc.), Alexa Fluor dyes, Atto dyes, photoswtichable dyes, photoactivatable dyes, fluorescent dyes, metal nanoparticles, semiconductor nanoparticles or "quantum dots", fluorescent proteins such as GFP (Green Fluorescent Protein), or photoactivabale fluorescent proteins, such as PAGFP, PSCFP, PSCFP2, Dendra, Dendra2, EosFP, tdEos, mEos2, mEos3, PAmCherry, PAtagRFP, mMaple, mMaple2, and mMaple3. Other suitable signaling entities are known to those of ordinary skill in the art. See, e.g., U.S. Pat. No. 7,838,302 or U.S. Pat. Apl. Ser. No. 61/979,436, each incorporated herein by reference in its entirety.

In one set of embodiments, the signaling entity may be attached to an oligonucleotide sequence via a bond that can be cleaved to release the signaling entity. In one set of embodiments, a fluorophore may be conjugated to an oligonucleotide via a cleavable bond, such as a photocleavable bond. Non-limiting examples of photocleavable bonds include, but are not limited to, 1-(2-nitrophenyl)ethyl, 2-nitrobenzyl, biotin phosphoramidite, acrylic phosphoramidite, diethylaminocoumarin, 1-(4,5-dimethoxy-2-nitrophenyl)ethyl, cyclo-dodecyl (dimethoxy-2-nitrophenyl)ethyl, 4-aminomethyl-3-nitrobenzyl, (4-nitro-3-(1-chlorocarbonyloxyethyl)phenyl)methyl-S-acetylthioic acid ester, (4-nitro-3-(1-thlorocarbonyloxyethyl)phenyl)methyl-3-(2-pyridyldithiopropionic acid) ester, 3-(4,4'-dimethoxytrityl)-1-(2-nitrophenyl)-propane-1,3-diol-[2- cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-trifluoroacetylcaproamidomethyl)phenyl]-ethyl-[2-cyano-ethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-(4,4'-dimethoxytrityloxy)butyramidomethyl)phenyl]-ethyl-[2-cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-(N-(4,4'-dimethoxytrityl))-biotinamidocaproamido-methyl)phenyl]-ethyl-[2-cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, or similar linkers. In another set of embodiments, the fluorophore may be conjugated to an oligonucleotide via a disulfide bond. The disulfide bond may be cleaved by a variety of reducing agents such as, but not limited to, dithiothreitol, dithioerythritol, beta-mercaptoethanol, sodium borohydride, thioredoxin, glutaredoxin, trypsinogen, hydrazine, diisobutylaluminum hydride, oxalic acid, formic acid, ascorbic acid, phosphorous acid, tin chloride, glutathione, thioglycolate, 2,3-dimercaptopropanol, 2-mercaptoethylamine, 2-aminoethanol, tris(2-carboxyethyl)phosphine, bis(2-mercaptoethyl) sulfone, N,N'-dimethyl-N,N'-bis(mercaptoacetyl)hydrazine, 3-mercaptoproptionate, dimethylformamide, thiopropyl-agarose, tri-*n*-butylphosphine, cysteine, iron sulfate, sodium sulfite, phosphite, hypophosphite, phosphorothioate, or the like, and/or combinations of any of these. In another embodiment, the fluorophore may be conjugated to an oligonucleotide via one or more phosphorothioate modified nucleotides in which the sulfur modification replaces the bridging and/or non-bridging oxygen. The fluorophore may be

cleaved from the oligonucleotide, in certain embodiments, via addition of compounds such as but not limited to iodoethanol, iodine mixed in ethanol, silver nitrate, or mercury chloride. In yet another set of embodiments, the signaling entity may be chemically inactivated through reduction or oxidation. For example, in one embodiment, a chromophore such as Cy5 or Cy7 may be reduced using sodium borohydride to a stable, non-fluorescence state. In still another set of embodiments, a fluorophore may be conjugated to an oligonucleotide via an azo bond, and the azo bond may be cleaved with 2-[(2-N-arylamino)phenylazo]pyridine. In yet another set of embodiments, a fluorophore may be conjugated to an oligonucleotide via a suitable nucleic acid segment that can be cleaved upon suitable exposure to DNAse, e.g., an exodeoxyribonuclease or an endodeoxyribonuclease. Examples include, but are not limited to, deoxyribonuclease I or deoxyribonuclease II. In one set of embodiments, the cleavage may occur via a restriction endonuclease. Non-limiting examples of potentially suitable restriction endonucleases include BamHI, BsrI, NotI, XmaI, PspAI, DpnI, MboI, MnlI, Eco57I, Ksp632I, DraIII, AhaII, SmaI, MluI, HpaI, ApaI, BclI, BstEII, TaqI, EcoRI, SacI, HindII, HaeII, DraII, Tsp509I, Sau3AI, PacI, etc. Over 3000 restriction enzymes have been studied in detail, and more than 600 of these are available commercially. In yet another set of embodiments, a fluorophore may be conjugated to biotin, and the oligonucleotide conjugated to avidin or streptavidin. An interaction between biotin and avidin or streptavidin allows the fluorophore to be conjugated to the oligonucleotide, while sufficient exposure to an excess of addition, free biotin could "outcompete" the linkage and thereby cause cleavage to occur. In addition, in another set of embodiments, the probes may be removed using corresponding "toe-hold-probes," which comprise the same sequence as the probe, as well as an extra number of bases of homology to the encoding probes (e.g., 1-20 extra bases, for example, 5 extra bases). These probes may remove the labeled readout probe through a strand-displacement interaction.

As used herein, the term "light" generally refers to electromagnetic radiation, having any suitable wavelength (or equivalently, frequency). For instance, in some embodiments, the light may include wavelengths in the optical or visual range (for example, having a wavelength of between about 400 nm and about 700 nm, i.e., "visible light"), infrared wavelengths (for example, having a wavelength of between about 300 micrometers and 700 nm), ultraviolet wavelengths (for example, having a wavelength of between about 400 nm and about 10 nm), or the like. In certain cases, as discussed in detail below, more than one entity may be used, i.e., entities that are chemically different or distinct, for example,

structurally. However, in other cases, the entities may be chemically identical or at least substantially chemically identical.

In one set of embodiments, the signaling entity is "switchable," i.e., the entity can be switched between two or more states, at least one of which emits light having a desired wavelength. In the other state(s), the entity may emit no light, or emit light at a different wavelength. For instance, an entity may be "activated" to a first state able to produce light having a desired wavelength, and "deactivated" to a second state not able to emit light of the same wavelength. An entity is "photoactivatable" if it can be activated by incident light of a suitable wavelength. As a non-limiting example, Cy5, can be switched between a fluorescent and a dark state in a controlled and reversible manner by light of different wavelengths, i.e., 633 nm (or 642nm, 647nm, 656 nm) red light can switch or deactivate Cy5 to a stable dark state, while 405 nm green light can switch or activate the Cy5 back to the fluorescent state. In some cases, the entity can be reversibly switched between the two or more states, e.g., upon exposure to the proper stimuli. For example, a first stimuli (e.g., a first wavelength of light) may be used to activate the switchable entity, while a second stimuli (e.g., a second wavelength of light) may be used to deactivate the switchable entity, for instance, to a non-emitting state. Any suitable method may be used to activate the entity. For example, in one embodiment, incident light of a suitable wavelength may be used to activate the entity to emit light, i.e., the entity is "photoswitchable." Thus, the photoswitchable entity can be switched between different light-emitting or non-emitting states by incident light, e.g., of different wavelengths. The light may be monochromatic (e.g., produced using a laser) or polychromatic. In another embodiment, the entity may be activated upon stimulation by electric field and/or magnetic field. In other embodiments, the entity may be activated upon exposure to a suitable chemical environment, e.g., by adjusting the pH, or inducing a reversible chemical reaction involving the entity, etc. Similarly, any suitable method may be used to deactivate the entity, and the methods of activating and deactivating the entity need not be the same. For instance, the entity may be deactivated upon exposure to incident light of a suitable wavelength, or the entity may be deactivated by waiting a sufficient time.

Typically, a "switchable" entity can be identified by one of ordinary skill in the art by determining conditions under which an entity in a first state can emit light when exposed to an excitation wavelength, switching the entity from the first state to the second state, e.g., upon exposure to light of a switching wavelength, then showing that the entity, while in the second state can no longer emit light (or emits light at a much reduced intensity) when exposed to the excitation wavelength.

In one set of embodiments, as discussed, a switchable entity may be switched upon exposure to light. In some cases, the light used to activate the switchable entity may come from an external source, e.g., a light source such as a laser light source, another light-emitting entity proximate the switchable entity, etc. The second, light emitting entity, in some cases, may be a fluorescent entity, and in certain embodiments, the second, light-emitting entity may itself also be a switchable entity.

In some embodiments, the switchable entity includes a first, light-emitting portion (e.g., a fluorophore), and a second portion that activates or "switches" the first portion. For example, upon exposure to light, the second portion of the switchable entity may activate the first portion, causing the first portion to emit light. Examples of activator portions include, but are not limited to, Alexa Fluor 405 (Invitrogen), Alexa Fluor 488 (Invitrogen), Cy2 (GE Healthcare), Cy3 (GE Healthcare), Cy3B (GE Healthcare), Cy3.5 (GE Healthcare), or other suitable dyes. Examples of light-emitting portions include, but are not limited to, Cy5, Cy5.5 (GE Healthcare), Cy7 (GE Healthcare), Alexa Fluor 647 (Invitrogen), Alexa Fluor 680 (Invitrogen), Alexa Fluor 700 (Invitrogen), Alexa Fluor 750 (Invitrogen), Alexa Fluor 790 (Invitrogen), DiD, DiR, YOYO-3 (Invitrogen), YO-PRO-3 (Invitrogen), TOT-3 (Invitrogen), TO-PRO-3 (Invitrogen) or other suitable dyes. These may linked together, e.g., covalently, for example, directly, or through a linker, e.g., forming compounds such as, but not limited to, Cy5-Alexa Fluor 405, Cy5-Alexa Fluor 488, Cy5-Cy2, Cy5-Cy3, Cy5-Cy3.5, Cy5.5-Alexa Fluor 405, Cy5.5-Alexa Fluor 488, Cy5.5-Cy2, Cy5.5-Cy3, Cy5.5-Cy3.5, Cy7-Alexa Fluor 405, Cy7-Alexa Fluor 488, Cy7-Cy2, Cy7-Cy3, Cy7-Cy3.5, Alexa Fluor 647-Alexa Fluor 405, Alexa Fluor 647-Alexa Fluor 488, Alexa Fluor 647-Cy2, Alexa Fluor 647-Cy3, Alexa Fluor 647-Cy3.5, Alexa Fluor 750-Alexa Fluor 405, Alexa Fluor 750-Alexa Fluor 488, Alexa Fluor 750-Cy2, Alexa Fluor 750-Cy3, or Alexa Fluor 750-Cy3.5. Those of ordinary skill in the art will be aware of the structures of these and other compounds, many of which are available commercially. The portions may be linked via a covalent bond, or by a linker, such as those described in detail below. Other light-emitting or activator portions may include portions having two quaternized nitrogen atoms joined by a polymethine chain, where each nitrogen is independently part of a heteroaromatic moiety, such as pyrrole, imidazole, thiazole, pyridine, quinoine, indole, benzothiazole, etc., or part of a nonaromatic amine. In some cases, there may be 5, 6, 7, 8, 9, or more carbon atoms between the two nitrogen atoms.

In certain cases, the light-emitting portion and the activator portions, when isolated from each other, may each be fluorophores, i.e., entities that can emit light of a certain,

emission wavelength when exposed to a stimulus, for example, an excitation wavelength. However, when a switchable entity is formed that comprises the first fluorophore and the second fluorophore, the first fluorophore forms a first, light-emitting portion and the second fluorophore forms an activator portion that switches that activates or "switches" the first portion in response to a stimulus. For example, the switchable entity may comprise a first fluorophore directly bonded to the second fluorophore, or the first and second entity may be connected via a linker or a common entity. Whether a pair of light-emitting portion and activator portion produces a suitable switchable entity can be tested by methods known to those of ordinary skills in the art. For example, light of various wavelength can be used to stimulate the pair and emission light from the light-emitting portion can be measured to determined wither the pair makes a suitable switch.

As a non-limiting example, Cy3 and Cy5 may be linked together to form such an entity. In this example, Cy3 is an activator portion that is able to activate Cy5, the light-emission portion. Thus, light at or near the absorption maximum (e.g., near 532 nm light for Cy3) of the activation or second portion of the entity may cause that portion to activate the first, light-emitting portion, thereby causing the first portion to emit light (e.g., near 647 nm for Cy5). See, e.g., U.S. Pat. No. 7,838,302, incorporated herein by reference in its entirety. In some cases, the first, light-emitting portion can subsequently be deactivated by any suitable technique (e.g., by directing 647 nm red light to the Cy5 portion of the molecule).

Other non-limiting examples of potentially suitable activator portions include 1,5 IAEDANS, 1,8-ANS, 4-Methylumbelliferone, 5-carboxy-2,7-dichlorofluorescein, 5-Carboxyfluorescein (5-FAM), 5-Carboxynapthofluorescein, 5-Carboxytetramethylrhodamine (5-TAMRA), 5-FAM (5-Carboxyfluorescein), 5-HAT (Hydroxy Tryptamine), 5-Hydroxy Tryptamine (HAT), 5-ROX (carboxy-X-rhodamine), 5-TAMRA (5-Carboxytetramethylrhodamine), 6-Carboxyrhodamine 6G, 6-CR 6G, 6-JOE, 7-Amino-4-methylcoumarin, 7-Aminoactinomycin D (7-AAD), 7-Hydroxy-4-methylcoumarin, 9-Amino-6-chloro-2-methoxyacridine, ABQ, Acid Fuchsin, ACMA (9-Amino-6-chloro-2-methoxyacridine), Acridine Orange, Acridine Red, Acridine Yellow, Acriflavin, Acriflavin Feulgen SITSA, Alexa Fluor 350, Alexa Fluor 405, Alexa Fluor 430, Alexa Fluor 488, Alexa Fluor 500, Alexa Fluor 514, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 555, Alexa Fluor 568, Alexa Fluor 594, Alexa Fluor 610, Alexa Fluor 633, Alexa Fluor 635, Alizarin Complexon, Alizarin Red, AMC, AMCA-S, AMCA (Aminomethylcoumarin), AMCA-X, Aminoactinomycin D, Aminocoumarin, Aminomethylcoumarin (AMCA), Anilin Blue, Anthrocyl stearate, APTRA-BTC, APTS, Astrazon Brilliant Red 4G, Astrazon Orange R,

Astrazon Red 6B, Astrazon Yellow 7 GLL, Atabrine, ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 520, ATTO 532, ATTO 550, ATTO 565, ATTO 590, ATTO 594, ATTO 610, ATTO 611X, ATTO 620, ATTO 633, ATTO 635, ATTO 647, ATTO 647N, ATTO 655, ATTO 680, ATTO 700, ATTO 725, ATTO 740, ATTO-TAG CBQCA, ATTO-TAG FQ, Auramine, Aurophosphine G, Aurophosphine, BAO 9 (Bisaminophenyloxadiazole), BCECF (high pH), BCECF (low pH), Berberine Sulphate, Bimane, Bisbenzamide, Bisbenzimide (Hoechst), bis-BTC, Blancophor FFG, Blancophor SV, BOBO -1, BOBO -3, Bodipy 492/515, Bodipy 493/503, Bodipy 500/510, Bodipy 505/515, Bodipy 530/550, Bodipy 542/563, Bodipy 558/568, Bodipy 564/570, Bodipy 576/589, Bodipy 581/591, Bodipy 630/650-X, Bodipy 650/665-X, Bodipy 665/676, Bodipy Fl, Bodipy FL ATP, Bodipy Fl-Ceramide, Bodipy R6G, Bodipy TMR, Bodipy TMR-X conjugate, Bodipy TMR-X, SE, Bodipy TR, Bodipy TR ATP, Bodipy TR-X SE, BO-PRO -1, BO-PRO -3, Brilliant Sulphoflavin FF, BTC, BTC-5N, Calcein, Calcein Blue, Calcium Crimson, Calcium Green, Calcium Green-1 Ca^{2+} Dye, Calcium Green-2 Ca^{2+}, Calcium Green-5N Ca^{2+}, Calcium Green-C18 Ca^{2+}, Calcium Orange, Calcofluor White, Carboxy-X-rhodamine (5-ROX), Cascade Blue, Cascade Yellow, Catecholamine, CCF2 (GeneBlazer), CFDA, Chromomycin A, Chromomycin A, CL-NERF, CMFDA, Coumarin Phalloidin, CPM Methylcoumarin, CTC, CTC Formazan, Cy2, Cy3.1 8, Cy3.5, Cy3, Cy5.1 8, cyclic AMP Fluorosensor (FiCRhR), Dabcyl, Dansyl, Dansyl Amine, Dansyl Cadaverine, Dansyl Chloride, Dansyl DHPE, Dansyl fluoride, DAPI, Dapoxyl, Dapoxyl 2, Dapoxyl 3' DCFDA, DCFH (Dichlorodihydrofluorescein Diacetate), DDAO, DHR (Dihydorhodamine 123), Di-4-ANEPPS, Di-8-ANEPPS (non-ratio), DiA (4-Di-16-ASP), Dichlorodihydrofluorescein Diacetate (DCFH), DiD - Lipophilic Tracer, DiD (DiIC18(5)), DIDS, Dihydorhodamine 123 (DHR), DiI (DiIC18(3)), Dinitrophenol, DiO (DiOC18(3)), DiR, DiR (DiIC18(7)), DM-NERF (high pH), DNP, Dopamine, DTAF, DY-630-NHS, DY-635-NHS, DyLight 405, DyLight 488, DyLight 549, DyLight 633, DyLight 649, DyLight 680, DyLight 800, ELF 97, Eosin, Erythrosin, Erythrosin ITC, Ethidium Bromide, Ethidium homodimer -1 (EthD-1), Euchrysin, EukoLight, Europium (III) chloride, Fast Blue, FDA, Feulgen (Pararosaniline), FIF (Formaldehyd Induced Fluorescence), FITC, Flazo Orange, Fluo-3, Fluo-4, Fluorescein (FITC), Fluorescein Diacetate, Fluoro-Emerald, Fluoro-Gold (Hydroxystilbamidine), Fluor-Ruby, FluorX, FM 1-43, FM 4-46, Fura Red (high pH), Fura Red/Fluo-3, Fura-2, Fura-2/BCECF, Genacryl Brilliant Red B, Genacryl Brilliant Yellow 10GF, Genacryl Pink 3G, Genacryl Yellow 5GF, GeneBlazer (CCF2), Gloxalic Acid, Granular blue, Haematoporphyrin, Hoechst 33258, Hoechst 33342, Hoechst 34580, HPTS,

Hydroxycoumarin, Hydroxystilbamidine (FluoroGold), Hydroxytryptamine, Indo-1, high calcium, Indo-1, low calcium, Indodicarbocyanine (DiD), Indotricarbocyanine (DiR), Intrawhite Cf, JC-1, JO-JO-1, JO-PRO-1, LaserPro, Laurodan, LDS 751 (DNA), LDS 751 (RNA), Leucophor PAF, Leucophor SF, Leucophor WS, Lissamine Rhodamine, Lissamine Rhodamine B, Calcein/Ethidium homodimer, LOLO-1, LO-PRO-1, Lucifer Yellow, Lyso Tracker Blue, Lyso Tracker Blue-White, Lyso Tracker Green, Lyso Tracker Red, Lyso Tracker Yellow, LysoSensor Blue, LysoSensor Green, LysoSensor Yellow/Blue, Mag Green, Magdala Red (Phloxin B), Mag-Fura Red, Mag-Fura-2, Mag-Fura-5, Mag-Indo-1, Magnesium Green, Magnesium Orange, Malachite Green, Marina Blue, Maxilon Brilliant Flavin 10 GFF, Maxilon Brilliant Flavin 8 GFF, Merocyanin, Methoxycoumarin, Mitotracker Green FM, Mitotracker Orange, Mitotracker Red, Mitramycin, Monobromobimane, Monobromobimane (mBBr-GSH), Monochlorobimane, MPS (Methyl Green Pyronine Stilbene), NBD, NBD Amine, Nile Red, Nitrobenzoxadidole, Noradrenaline, Nuclear Fast Red, Nuclear Yellow, Nylosan Brilliant Iavin E8G, Oregon Green, Oregon Green 488-X, Oregon Green, Oregon Green 488, Oregon Green 500, Oregon Green 514, Pacific Blue, Pararosaniline (Feulgen), PBFI, Phloxin B (Magdala Red), Phorwite AR, Phorwite BKL, Phorwite Rev, Phorwite RPA, Phosphine 3R, PKH26 (Sigma), PKH67, PMIA, Pontochrome Blue Black, POPO-1, POPO-3, PO-PRO-1, PO-PRO-3, Primuline, Procion Yellow, Propidium Iodid (PI), PyMPO, Pyrene, Pyronine, Pyronine B, Pyrozal Brilliant Flavin 7GF, QSY 7, Quinacrine Mustard, Resorufin, RH 414, Rhod-2, Rhodamine, Rhodamine 110, Rhodamine 123, Rhodamine 5 GLD, Rhodamine 6G, Rhodamine B, Rhodamine B 200, Rhodamine B extra, Rhodamine BB, Rhodamine BG, Rhodamine Green, Rhodamine Phallicidine, Rhodamine Phalloidine, Rhodamine Red, Rhodamine WT, Rose Bengal, S65A, S65C, S65L, S65T, SBFI, Serotonin, Sevron Brilliant Red 2B, Sevron Brilliant Red 4G, Sevron Brilliant Red B, Sevron Orange, Sevron Yellow L, SITS, SITS (Primuline), SITS (Stilbene Isothiosulphonic Acid), SNAFL calcein, SNAFL-1, SNAFL-2, SNARF calcein, SNARF1, Sodium Green, SpectrumAqua, SpectrumGreen, SpectrumOrange, Spectrum Red, SPQ (6-methoxy-N-(3-sulfopropyl)quinolinium), Stilbene, Sulphorhodamine B can C, Sulphorhodamine Extra, SYTO 11, SYTO 12, SYTO 13, SYTO 14, SYTO 15, SYTO 16, SYTO 17, SYTO 18, SYTO 20, SYTO 21, SYTO 22, SYTO 23, SYTO 24, SYTO 25, SYTO 40, SYTO 41, SYTO 42, SYTO 43, SYTO 44, SYTO 45, SYTO 59, SYTO 60, SYTO 61, SYTO 62, SYTO 63, SYTO 64, SYTO 80, SYTO 81, SYTO 82, SYTO 83, SYTO 84, SYTO 85, SYTOX Blue, SYTOX Green, SYTOX Orange, Tetracycline, Tetramethylrhodamine (TAMRA), Texas Red, Texas Red-X conjugate, Thiadicarbocyanine (DiSC3), Thiazine Red

R, Thiazole Orange, Thioflavin 5, Thioflavin S, Thioflavin TCN, Thiolyte, Thiozole Orange, Tinopol CBS (Calcofluor White), TMR, TO-PRO-1, TO-PRO-3, TO-PRO-5, TOTO-1, TOTO-3, TRITC (tetramethylrodamine isothiocyanate), True Blue, TruRed, Ultralite, Uranine B, Uvitex SFC, WW 781, X-Rhodamine, XRITC, Xylene Orange, Y66F, Y66H, Y66W, YO-PRO-1, YO-PRO-3, YOYO-1, YOYO-3, SYBR Green, Thiazole orange (interchelating dyes), or combinations thereof.

Another aspect of the invention is directed to a computer-implemented method. For instance, a computer and/or an automated system may be provided that is able to automatically and/or repetitively perform any of the methods described herein. As used herein, "automated" devices refer to devices that are able to operate without human direction, i.e., an automated device can perform a function during a period of time after any human has finished taking any action to promote the function, e.g. by entering instructions into a computer to start the process. Typically, automated equipment can perform repetitive functions after this point in time. The processing steps may also be recorded onto a machine-readable medium in some cases.

For example, in some cases, a computer may be used to control imaging of the sample, e.g., using fluorescence microscopy, STORM or other super-resolution techniques such as those described herein. In some cases, the computer may also control operations such as drift correction, physical registration, hybridization and cluster alignment in image analysis, cluster decoding (e.g., fluorescent cluster decoding), error detection or correction (e.g., as discussed herein), noise reduction, identification of foreground features from background features (such as noise or debris in images), or the like. As an example, the computer may be used to control activation and/or excitation of signaling entities within the sample, and/or the acquisition of images of the signaling entities. In one set of embodiments, a sample may be excited using light having various wavelengths and/or intensities, and the sequence of the wavelengths of light used to excite the sample may be correlated, using a computer, to the images acquired of the sample containing the signaling entities. For instance, the computer may apply light having various wavelengths and/or intensities to a sample to yield different average numbers of signaling entities in each region of interest (e.g., one activated entity per location, two activated entities per location, etc.). In some cases, this information may be used to construct an image and/or determine the locations of the signaling entities, in some cases at high resolutions, as noted above.

In some aspects, the sample is positioned on a microscope. In some cases, the microscope may contain one or more channels, such as microfluidic channels, to direct or

control fluid to or from the sample. For instance, in one embodiment, nucleic acid probes such as those discussed herein may be introduced and/or removed from the sample by flowing fluid through one or more channels to or from the sample. In some cases, there may also be one or more chambers or reservoirs for holding fluid, e.g., in fluidic communication with the channel, and/or with the sample. Those of ordinary skill in the art will be familiar with channels, including microfluidic channels, for moving fluid to or from a sample.

As used herein, "microfluidic," "microscopic," "microscale," the "micro-" prefix (for example, as in "microchannel"), and the like generally refers to elements or articles having widths or diameters of less than about 1 mm, and less than about 100 microns (micrometers) in some cases. In some embodiments, larger channels may be used instead of, or in conjunction with, microfluidic channels for any of the embodiments discussed herein. For examples, channels having widths or diameters of less than about 10 mm, less than about 9 mm, less than about 8 mm, less than about 7 mm, less than about 6 mm, less than about 5 mm, less than about 4 mm, less than about 3 mm, or less than about 2 mm may be used in certain instances. In some cases, the element or article includes a channel through which a fluid can flow. In all embodiments, specified widths can be a smallest width (i.e. a width as specified where, at that location, the article can have a larger width in a different dimension), or a largest width (i.e. where, at that location, the article has a width that is no wider than as specified, but can have a length that is greater). Thus, for instance, the microfluidic channel may have an average cross-sectional dimension (e.g., perpendicular to the direction of flow of fluid in the microfluidic channel) of less than about 1 mm, less than about 500 microns, less than about 300 microns, or less than about 100 microns. In some cases, the microfluidic channel may have an average diameter of less than about 60 microns, less than about 50 microns, less than about 40 microns, less than about 30 microns, less than about 25 microns, less than about 10 microns, less than about 5 microns, less than about 3 microns, or less than about 1 micron.

A "channel," as used herein, means a feature on or in an article (e.g., a substrate) that at least partially directs the flow of a fluid. In some cases, the channel may be formed, at least in part, by a single component, e.g. an etched substrate or molded unit. The channel can have any cross-sectional shape, for example, circular, oval, triangular, irregular, square or rectangular (having any aspect ratio), or the like, and can be covered or uncovered (i.e., open to the external environment surrounding the channel). In embodiments where the channel is completely covered, at least one portion of the channel can have a cross-section that is

completely enclosed, and/or the entire channel may be completely enclosed along its entire length with the exception of its inlet and outlet.

A channel may have any aspect ratio, e.g., an aspect ratio (length to average cross-sectional dimension) of at least about 2:1, more typically at least about 3:1, at least about 5:1, at least about 10:1, etc. As used herein, a "cross-sectional dimension," in reference to a fluidic or microfluidic channel, is measured in a direction generally perpendicular to fluid flow within the channel. A channel generally will include characteristics that facilitate control over fluid transport, e.g., structural characteristics and/or physical or chemical characteristics (hydrophobicity vs. hydrophilicity) and/or other characteristics that can exert a force (e.g., a containing force) on a fluid. The fluid within the channel may partially or completely fill the channel. In some cases the fluid may be held or confined within the channel or a portion of the channel in some fashion, for example, using surface tension (e.g., such that the fluid is held within the channel within a meniscus, such as a concave or convex meniscus). In an article or substrate, some (or all) of the channels may be of a particular size or less, for example, having a largest dimension perpendicular to fluid flow of less than about 5 mm, less than about 2 mm, less than about 1 mm, less than about 500 microns, less than about 200 microns, less than about 100 microns, less than about 60 microns, less than about 50 microns, less than about 40 microns, less than about 30 microns, less than about 25 microns, less than about 10 microns, less than about 3 microns, less than about 1 micron, less than about 300 nm, less than about 100 nm, less than about 30 nm, or less than about 10 nm or less in some cases. In one embodiment, the channel is a capillary.

A variety of materials and methods, according to certain aspects of the invention, can be used to form devices or components containing microfluidic channels, chambers, etc. For example, various devices or components can be formed from solid materials, in which the channels can be formed via micromachining, film deposition processes such as spin coating and chemical vapor deposition, physical vapor deposition, laser fabrication, photolithographic techniques, etching methods including wet chemical or plasma processes, electrodeposition, and the like. See, for example, *Scientific American*, 248:44-55, 1983 (Angell, *et al*).

In one set of embodiments, various structures or components can be formed of a polymer, for example, an elastomeric polymer such as polydimethylsiloxane ("PDMS"), polytetrafluoroethylene ("PTFE" or Teflon®), or the like. For instance, according to one embodiment, a channel such as a microfluidic channel may be implemented by fabricating the fluidic system separately using PDMS or other soft lithography techniques (details of soft lithography techniques suitable for this embodiment are discussed in the references entitled

"Soft Lithography," by Younan Xia and George M. Whitesides, published in the *Annual Review of Material Science*, 1998, Vol. 28, pages 153-184, and "Soft Lithography in Biology and Biochemistry," by George M. Whitesides, Emanuele Ostuni, Shuichi Takayama, Xingyu Jiang and Donald E. Ingber, published in the *Annual Review of Biomedical Engineering*, 2001, Vol. 3, pages 335-373; each of these references is incorporated herein by reference).

Other examples of potentially suitable polymers include, but are not limited to, polyethylene terephthalate (PET), polyacrylate, polymethacrylate, polycarbonate, polystyrene, polyethylene, polypropylene, polyvinylchloride, cyclic olefin copolymer (COC), polytetrafluoroethylene, a fluorinated polymer, a silicone such as polydimethylsiloxane, polyvinylidene chloride, bis-benzocyclobutene ("BCB"), a polyimide, a fluorinated derivative of a polyimide, or the like. Combinations, copolymers, or blends involving polymers including those described above are also envisioned. The device may also be formed from composite materials, for example, a composite of a polymer and a semiconductor material.

In some embodiments, various microfluidic structures or components of the device are fabricated from polymeric and/or flexible and/or elastomeric materials, and can be conveniently formed of a hardenable fluid, facilitating fabrication via molding (e.g. replica molding, injection molding, cast molding, etc.). The hardenable fluid can be essentially any fluid that can be induced to solidify, or that spontaneously solidifies, into a solid capable of containing and/or transporting fluids contemplated for use in and with the fluidic network. In one embodiment, the hardenable fluid comprises a polymeric liquid or a liquid polymeric precursor (i.e. a "prepolymer"). Suitable polymeric liquids can include, for example, thermoplastic polymers, thermoset polymers, waxes, metals, or mixtures or composites thereof heated above their melting point. As another example, a suitable polymeric liquid may include a solution of one or more polymers in a suitable solvent, which solution forms a solid polymeric material upon removal of the solvent, for example, by evaporation. Such polymeric materials, which can be solidified from, for example, a melt state or by solvent evaporation, are well known to those of ordinary skill in the art. A variety of polymeric materials, many of which are elastomeric, are suitable, and are also suitable for forming molds or mold masters, for embodiments where one or both of the mold masters is composed of an elastomeric material. A non-limiting list of examples of such polymers includes polymers of the general classes of silicone polymers, epoxy polymers, and acrylate polymers. Epoxy polymers are characterized by the presence of a three-membered cyclic ether group commonly referred to as an epoxy group, 1,2-epoxide, or oxirane. For example, diglycidyl

ethers of bisphenol A can be used, in addition to compounds based on aromatic amine, triazine, and cycloaliphatic backbones. Another example includes the well-known Novolac polymers. Non-limiting examples of silicone elastomers suitable for use according to the invention include those formed from precursors including the chlorosilanes such as methylchlorosilanes, ethylchlorosilanes, phenylchlorosilanes, etc.

Silicone polymers are used in certain embodiments, for example, the silicone elastomer polydimethylsiloxane. Non-limiting examples of PDMS polymers include those sold under the trademark Sylgard by Dow Chemical Co., Midland, MI, and particularly Sylgard 182, Sylgard 184, and Sylgard 186. Silicone polymers including PDMS have several beneficial properties simplifying fabrication of various structures of the invention. For instance, such materials are inexpensive, readily available, and can be solidified from a prepolymeric liquid via curing with heat. For example, PDMSs are typically curable by exposure of the prepolymeric liquid to temperatures of about, for example, about 65 °C to about 75 °C for exposure times of, for example, at least about an hour. Also, silicone polymers, such as PDMS, can be elastomeric and thus may be useful for forming very small features with relatively high aspect ratios, necessary in certain embodiments of the invention. Flexible (e.g., elastomeric) molds or masters can be advantageous in this regard.

One advantage of forming structures such as microfluidic structures or channels from silicone polymers, such as PDMS, is the ability of such polymers to be oxidized, for example by exposure to an oxygen-containing plasma such as an air plasma, so that the oxidized structures contain, at their surface, chemical groups capable of cross-linking to other oxidized silicone polymer surfaces or to the oxidized surfaces of a variety of other polymeric and non-polymeric materials. Thus, structures can be fabricated and then oxidized and essentially irreversibly sealed to other silicone polymer surfaces, or to the surfaces of other substrates reactive with the oxidized silicone polymer surfaces, without the need for separate adhesives or other sealing means. In most cases, sealing can be completed simply by contacting an oxidized silicone surface to another surface without the need to apply auxiliary pressure to form the seal. That is, the pre-oxidized silicone surface acts as a contact adhesive against suitable mating surfaces. Specifically, in addition to being irreversibly sealable to itself, oxidized silicone such as oxidized PDMS can also be sealed irreversibly to a range of oxidized materials other than itself including, for example, glass, silicon, silicon oxide, quartz, silicon nitride, polyethylene, polystyrene, glassy carbon, and epoxy polymers, which have been oxidized in a similar fashion to the PDMS surface (for example, via exposure to an oxygen-containing plasma). Oxidation and sealing methods useful in the context of the

present invention, as well as overall molding techniques, are described in the art, for example, in an article entitled "Rapid Prototyping of Microfluidic Systems and Polydimethylsiloxane," *Anal. Chem.*, 70:474-480, 1998 (Duffy *et al.*), incorporated herein by reference.

The following documents are each incorporated herein by reference in their entireties: U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*; and Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*

In addition, incorporated herein by reference in their entireties are U.S. Provisional Patent Application Serial No. 62/031,062, filed July 30, 2014, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/050,636, filed September 15, 2014, entitled "Probe Library Construction," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/142,653, filed April 3, 2015, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; and a PCT application filed on even date herewith, entitled "Probe Library Construction," by Zhuang, *et al.*

The following examples are intended to illustrate certain embodiments of the present invention, but do not exemplify the full scope of the invention.

EXAMPLE 1

The example presents a platform to enable the simultaneous detection of the number and spatial organization of thousands of distinct mRNAs within single cells with high efficiency and low error-rate using a novel form of highly multiplexed fluorescence *in situ* hybridization (FISH). This example accomplishes these measurements by integrating and innovating methods for massively parallel probe synthesis, super-resolution imaging, and self-correcting error-checking codes.

Here, these examples present methods for the simultaneous detection of some or all of the thousands of unique RNAs expressed in a cell. This approach not only promises to revolutionize the throughput of the already effective single-molecule FISH (smFISH) approach, but also allows researchers to benefit from the hypothesis free discovery approach which has made other whole-genome systems approaches to biology so effective. For example, this whole genome approach may allow researchers to discover RNAs whose expression levels and/or subcellular localization patterns change under certain conditions of interest, such as disease states, without knowing, *a priori*, which mRNA will change in

abundance or localization. Simultaneous measurements of hundreds of genes within a single cell also allow for the identification of correlations between genes in expression and localization patterns in some cases.

This can be achieved using methods for highly multiplexed smFISH via the sequential hybridization of orthogonal detection probes and super-resolution imaging, reducing the cost of probe synthesis, and the development of a highly automated system to minimize demands on the user, as discussed herein. This provides an integrated platform to handle the bioinformatics of probe design, the mathematics of error-correcting codes, the complexity of image registration and analysis, and the cumbersome fluid handling through a simple suite of user friendly interfaces. This integration allows easy operation with limited user training and facilitates the rapid collection of data.

This example illustrates: (1) computational design of "codewords" to attach to all RNA targets in the cell that will allow unique identification of each RNA with some degree of experimental error tolerance, (2) translation of these codewords into nucleotide sequences and synthesis of the required single-stranded (ss) oligonucleotide (e.g. ssDNA) probes, (3) sample fixation and hybridization of these probes to the RNA targets *in situ*, (4) read-out of these codewords via successive rounds of hybridization of distinct fluorescent probes imaged with conventional fluorescence microscopy or super-resolution fluorescence microscopy, and (5) automated decoding of measured codewords combined with computational error correction to uniquely and robustly identify individual mRNAs.

In the first step, a "codeword" is assigned to every RNA that is to be labeled. In a typical design these may be strings of N binary letters or positions. Codewords may be chosen from the same wide range of existing error tolerant or error-correcting encoding schemes developed for digital storage and communication, e.g., using Hamming codes or the like. For example, actin-RNA may be assigned the binary codeword 11001010. Each codeword may be unique and separated from the other codewords by a Hamming distance, h, which measures the number of letters or positions that must be incorrectly read for one codeword to be misinterpreted as a different one. A Hamming distance greater than 1 between all codewords allows for some measurement errors to be detected—since simple errors would produce codewords that are not used to encode RNAs. For a Hamming distance larger than 2, it is also possible to correct some errors, as codewords with one error will be closest in Hamming distance to a single, unique codeword. The total number of different RNAs to be detected from the transcriptome and the amount of error correction desired

determines the length of the codewords. Information theory provides several efficient algorithms for assembling error-correcting binary codebooks.

In the second step, this encoding scheme is translated into a set of oligonucleotide (e.g. DNA) probe sequences, which can be called primary probes or encoding probes, each of which not only targets a probe to the RNA of interest but also encodes the unique binary codeword within a set of secondary binding sites (Fig. 1). For example, first designed may be primary binding sequences for each targeted mRNA. These sequences are "target sequences" that are comprised of complementary nucleotide sequences to their target RNAs computationally selected to satisfy a stringent set of hybridization conditions, including uniqueness in the target genome. To improve the efficiency of hybridization to individual mRNAs, multiple primary target sequences are designed for each individual RNA. Then each position within the set of codewords is assigned a unique oligonucleotide (e.g. DNA) sequence, which is called a read sequence. These tags are designed as to have no interaction with endogenous mRNA sequences or each other. For instance, for all the value "1"s in a codeword of an individual mRNA, the corresponding read sequence is attached to the primary targeting sequences against that mRNA. In general, each probe will contain a target sequence and one or more read sequences. If the total length of the necessary read sequences and the primary target sequence exceeds synthesis capabilities, then subsets of the read sequences can be appended to distinct target sequences. For example, consider the potential codeword 11001010 for actin. Probe sequences for this RNA could contain the read sequences corresponding to positions 1, 2, 5, and 7 in the codeword attached to a variety of target sequences specific to actin. After all the sequences have been designed, the resulting complex set of unique custom oligonucleotide (e.g. DNA) sequences is manufactured and amplified using methods as described below.

In the third step, the resulting pool of DNA is hybridized, e.g., to fixed, permeabilized cells. In this process, individual probes may be attached to every RNA in the cell by hybridization of their corresponding target sequences with the RNA while the read sequences remain free to bind the appropriate secondary probes as discussed below.

In the fourth step—the read-out step—fluorescently labeled secondary nucleic acid probes (also called readout probes) are successively hybridized to the read sequences attached to the target sequences that binds to the mRNA targets in the above step. When simultaneously imaging a large number of different RNA species in cells, the density of labeled RNAs may exceed that at which each RNA can be resolved via conventional imaging methods. Thus, this may be performed using a super-resolution imaging method, for example

STORM (stochastic optical reconstruction microscopy), to resolve the labeled molecules. After each round of hybridization and imaging with the secondary probes, the fluorophores are quenched or otherwise inactivated either via chemical or optical techniques such as oxidation, chemically bleaching, photobleaching, stringent washing or enzymatic digestion, etc. The sample is then stained with the next secondary probe, and the cycle continues until all positions of the codewords have been read out. In the simplest incarnation, there will be one hybridization step for each position within the codeword, e.g. 8 hybridization steps for an 8-letter codeword (Fig. 1).

Fig. 1 shows schematic diagrams of this example. Fig. 1A shows that every position of the codewords is assigned a unique oligonucleotide sequence when this position has a value "1." All mRNA codewords are then translated into combinations of read sequences, which are attached to the targeting sequence. Fig. 1B shows various steps of the labeling scheme of this example. In the first step, all mRNAs (I - III) are tagged with multiple oligonucleotide (e.g. ssDNA) probes comprising a primary targeting sequence, which hybridizes to the RNA of interest, and a "tail" (i.e. containing read sequences) carrying the translated codeword, which does not interact with endogenous nucleotide sequences. In the next step, the first secondary probe is added, which can bind all probes whose tails have a read sequence corresponding to the value of "1" in the first position. The dyes on these secondary probes are imaged and bleached, then the next secondary probe is added to bind probes attached to mRNA which have a value of "1" in the second position of their assigned codeword, and so on.

In the final step, the microscopy images from each staining and imaging round are aligned, for example, computationally (e.g. using fiducial beads or other markers tracked during image acquisition), and the clusters of localizations resolved by conventional fluorescence microscopy or super-resolution imaging (e.g. STORM) from the different rounds are identified. These clusters of localizations arise from individual target mRNA molecules, and the hybridization rounds in which a spot was detected in a given cluster correspond to the "1" in the codeword for that mRNA. If there are no missed-detection events or false positive signal in the images, this codeword will perfectly match one of the expected codewords. Fig. 1 describes an example in which the codeword has three letters, i.e. three positions, and the three target mRNAs have codewords 110, 101, and 011 assigned to them. In real experimental examples, the codeword could contain more digits. For example, the mRNA for actin can be assigned the codeword, 11001010. In that case, detected clusters containing overlapping localization signals in the 1st, 2nd, 5th and 7th hybridization steps

(meaning the 1st, 2nd, 5th and 7th secondary probes bound to this site) can be identified as individual actin mRNA molecules, since the pattern of positive bindings matches the codewode of actin (11001010). In addition, if there are missed-detection events or false positive signals in the image data, these aberrations can be corrected by the implemented error-correction scheme. For example, clusters of localizations with a detected codeword that has only one digit discrepancy from 11001010 (such as 11000010 or 11101010) can also be identified as actin mRNA since all other valid codeword in this example differ from the detected pattern in two more positions.

EXAMPLE 2

This example describes another alternative approach that differs in several of the steps described above. This approach begins with the first step, construction of the codewords to the desired mRNA targets, as described above.

In the second step of this approach, nucleic acid probes are designed that bind uniquely to the mRNA targets of interest, as described above. However, instead of appending unique read sequences to these targeting sequences, unique pools or groups of probes are constructed from these target sequences. Each pool comprises all or a subset of the sequences that target all mRNAs which contain the same value at a given position in their codeword. For example, the first pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the first position of their codewords, e.g. 110 and 101 but not 011; the second pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the second position of their codewords, e.g. 110 and 011 but not 101; the third pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the third position of their codewords, e.g. 011 and 101 but not 110 (Fig. 1C). As another example, consider the potential codeword 11001010 for actin. Probes that target this mRNA would be included in pools 1, 2, 5, and 7 but not in pools 3, 4, 6, and 8. The same target to a given mRNA may or may not be included in pools. For example, a probe that targets the same region of actin may be included in pools 1, 2, 5, and 7 or any subset of these pools. After all of the pools have been designed, each complex set of unique, custom oligonucleotide sequences is manufactured and amplified using methods as described below.

In the third step of this approach, the first pool of probes is hybridized, e.g., to fixed, permeabilized cells. In this process, the fluorophores attached to each of the probes in this pool are bound to each of the targets of that pool. The binding of these probes is then determined by fluorescence microscopy. As described above, these images can be collected either via a range of methods including both conventional fluorescence imaging or super-

resolution imaging methods such as STORM. After a round of imaging, the probes from the first pool are inactivated or removed from the sample via the methods described above. This process is then repeated for each successive pool of probes until some or all of the pools have been applied to the sample and imaged such that all positions in the codewords have been read out. In the simplest incarnation, there will be one hybridization and imaging step for each position in the codeword, e.g. 3 rounds of hybridization and imaging for a codeword with 3 positions (Fig 1C) or 8 rounds of hybridization and imaging for a codeword with 8 positions.

The final step of this approach is identical to that described above.

EXAMPLE 3

In this example, 14 genes (PGK1, H3F3B, PKM, ENO1, GPI, EEF2, GNAS, HSPA8, GAPDH, CALM1, RHOA, PPIA, UBA52, and VCP) were encoded using a subset of the (8,4) SECDED code (Figs. 2A-2E). To determine the accuracy of these measurements, the measured abundances of these 14 mRNAs were compared to the abundances measured from bulk RNA-seq of A549 cells (published ENCODE data). Remarkably, it was found that there was excellent agreement between these two measurements, as the transcripts count measured using the sequential hybridization approach correlated with gene expression measured using RNA-seq with a Pearson correlation coefficient r of 0.75 (Fig. 2F). Gene expression from 3 other cells were also measured, and it was found that the gene expression of these 14 genes was highly correlated among the cells with a r of 0.96 (Fig. 2G).

Codebook Design. Each mRNA in the target set was assigned a binary codeword using a Single Error Correction Double Error Detection (SECDED) code. SECDED is an extended Hamming codebook with an additional parity bit. Briefly, Matlab's Communications System toolbox was used to generate SECDED codes of either 8 or 16 letters or positions. In both cases, only those codewords containing four 1s were used. These words were assigned at random to mRNAs in the target set. [0 1 0 1 1 1 0 0] is an example of the 8-letter codewords used (i.e., these codewords each contained four 1s and four 0s.) [0 1 0 1 1 1 0 0 0 0 0 0 0 0 0 0] is an example of the 16-letter codewords used (i.e., each codeword contained four 1s and twelve 0s). Not every codeword was necessarily assigned to an mRNA.

Computational Assembly of ssDNA Primary Probe Sequence. The number of primary nucleic acid probes used for hybridization with mRNA targets ranged from 200 to 2000 unique oligonucelotides, depending on the experiment. For example, to label 14 mRNAs with 28 oligos targeting each gene, 392 unique sequences were used. Large number

of oligos with unique sequences were purchased in a pool from LC Sciences or CustomArray. However, array synthesized oligos were in minute quantity that was insufficient for *in situ* hybridization. The protocol for their amplification are described below.

Each primary probe contained three components: flanking primer sequences to allow enzymatic amplification of probes, targeting sequence for *in situ* hybridization to mRNAs, and secondary tag sequence containing one or more read sequences for sequential readout of codewords.

The following is an example of a primary probe:

GTTGGCGACGAAAGCACTGCGATTGGAACCGTCCCAAGCGTTGCG
CTTAATGGATCATCAATTTTGTCTCACTACGACGGTCAATCGCGCTGCATACTTG
CGTCGGTCGGACAAACGAGG

(SEQ ID NO: 1)

The components are arranged in the following order: forward primer (not underlined), secondary read sequence 1 (underlined), mRNA targeting sequence (not underlined), secondary read sequence 2 (underlined), and reverse primer (not underlined). The secondary read sequences are the reverse complement of the corresponding secondary probes. Since only codewords that contained four '1's were used, the primary probes for each mRNA needed to contain 4 different secondary read sequences in this example. However, in order to reduce the overall length of the primary probes, the pool of targeting sequences for each mRNA target was split at random into two pools. Two secondary read sequences are attached each probe in one of the two pools and the other two secondary read sequences are attached the probes in the other pool. The design criteria for each component are described below.

Primer Design. Specific index primers were generated by a collection of 240,000 published sequences of orthogonal 25-bp long sequences. These sequences were trimmed to 20 bp, selected for a narrow 70 to 80 ^{o}C melting temperature, the absence of consecutive repeats of 3 or more base, and the presence of a GC clamp, i.e. one of the two 3' terminal bases must be G or C. To further improve specificity, these sequences were then screened against the human genome using BLAST+ (Camacho et al 2009), and primers with 14 or more contiguous bases of homology were eliminated. In a subsequent screening via BLAST+, primers that shared 11 or more contiguous bases or more than 5 bases at the 3' end of any other primer or the T7 promoter were also removed.

Secondary Probes Design. 30-bp long secondary probe sequences were created by concatenating fragments of the orthogonal primer set described above. These secondaries were then screened for orthogonality with other secondaries (no more than 11 basepairs of homology) and for potential off-target binding sites in the human genome (no more than 14 basepairs of homology). Secondary sequences used in this example are provided in Table 1.

Table 1

Bit	Secondary sequences	Sequence number
B1	CGCAACGCTTGGGACGGTTCCAATCGGATC	SEQ ID NO: 2
B2	CGAATGCTCTGGCCTCGAACGAACGATAGC	SEQ ID NO: 3
B3	ACAAATCCGACCAGATCGGACGATCATGGG	SEQ ID NO: 4
B4	CAAGTATGCAGCGCGATTGACCGTCTCGTT	SEQ ID NO: 5
B5	TGCGTCGTCTGGCTAGCACGGCACGCAAAT	SEQ ID NO: 6
B6	AAGTCGTACGCCGATGCGCAGCAATTCACT	SEQ ID NO: 7
B7	CGAAACATCGGCCACGGTCCCGTTGAACTT	SEQ ID NO: 8
B8	ACGAATCCACCGTCCAGCGCGTCAAACAGA	SEQ ID NO: 9
B9	CGCGAAATCCCCGTAACGAGCGTCCCTTGC	SEQ ID NO: 10
B10	GCATGAGTTGCCTGGCGTTGCGACGACTAA	SEQ ID NO: 11
B11	CCGTCGTCTCCGGTCCACCGTTGCGCTTAC	SEQ ID NO: 12
B12	GGCCAATGGCCCAGGTCCGTCACGCAATTT	SEQ ID NO: 13
B13	TTGATCGAATCGGAGCGTAGCGGAATCTGC	SEQ ID NO: 14
B14	CGCGCGGATCCGCTTGTCGGGAACGGATAC	SEQ ID NO: 15
B15	GCCTCGATTACGACGGATGTAATTCGGCCG	SEQ ID NO: 16
B16	GCCCGTATTCCCGCTTGCGAGTAGGGCAAT	SEQ ID NO: 17

mRNA Targeting Sequence Design. To determine the relative abundance of all the isoforms of all genes expressed in these cell lines, transcriptome profiling data from the ENCODE project for total RNA from A549 and IMR90 cells was processed using the publicly available software *cufflinks*, along with human genome annotations from gencode v18. Gene models corresponding to the most highly expressed isoform were used to build a sequence library in FASTA format recording the dominant isoform of every gene. Genes of interest were selected from this library. These genes were partitioned into 1kb segments, then

the software *OligoArray2.1* was used to generate primary probe sequences for the human transcriptome with the following constraints: 30-bp or 40-bp length, depending on the experiment; probe-target melting temperatures greater than 70 °C (variable parameter); no cross hybridization targets with melting temperatures greater than 72 °C (variable parameter); no predicted internal secondary structures with melting temperatures greater than 76 °C (variable parameter); and no single-nucleotide contiguous repeats of 6 or more bases. After OligoArray probe selection, all potential probes that mapped to a different gene were rejected while all potential probes with multiple alignments to the same gene were retained. A BLAST database was assembled from the FASTA library of all expressed genes to screen for probes' uniqueness. For each gene, 14 to 28 targeting sequences produced during the OligoArray processing were selected.

Probe Synthesis – index PCR. The template for specific probe sets were selected from the complex oligopool via limited-cycle PCR. Briefly, 0.5 to1 ng of the complex oligopool was combined with 0.5 micromolar of each primer. The forward primer matched the priming sequence for the desired subset while the reverse primer was a 5' concatenation of this sequence with a T7 promoter. To avoid the generation of G-quadruplets, which can be difficult to synthesize, the terminal Gs required in the T7 promoter were generated from Gs located at the 5' of the priming region where appropriate. All primers were synthesized by IDT. A 50 microliter reaction volume was amplified either using the KAPA real-time library amplification kit (KAPA Biosystems; KK2701) or via a homemade qPCR mix which included 0.8X EvaGreen (Biotum; 31000-T) and the hot-start Phusion polymerase (New England Biolabs; M0535S). Amplification was followed in real time using Agilent's MX300P or Biorad's CFX Connect. Individual samples were removed immediately before the plateau in amplification to minimize distortion of template abundance due to over-amplification. Individual templates were purified with columns according to the manufacturer's instructions (Zymo DNA Clean and Concentrator; D4003) and eluted in RNase-free deionized water.

Amplification via *in-vitro* transcription. The template was then amplified via *in vitro* transcription. Briefly, 0.5 to 1 microgram of template DNA was amplified into 100-200 microgram of RNA in a single 20-30 microliter reaction with a high yield RNA polymerase (New England Biolabs; E2040S). Reactions were supplemented with 1X RNase inhibitor (Promega RNasin; N2611). Amplification was typically run for 4 to 16 hours at 37 °C to maximize the yield. The RNA was not purified after the reaction and was either stored at -80 °C or immediately converted into DNA as described below.

Reverse Transcription. 1 to 2 nmol of fluorescently-labeled ssDNA probe was created from the above *in vitro* transcription reactions using the reverse transcriptase Maxima H- (Thermo Scientific; EP0751). This enzyme was used because of its higher processivity and temperature resistance, which allowed the conversion of large quantities of RNA into DNA within small volumes at temperatures that disfavor secondary structure formation. The unpurified RNA created above was supplemented with 1.6 mM of each dNTP, 1-2 nmol of fluorescently labeled forward primer, 300 units of Maxima H-, 60 units of RNasin, and a final 1X concentration of the Maxima RT buffer. The final 75 microliter volume was incubated at 50 °C for 60 minutes.

Strand Selection and Purification. The template RNA in the reaction above was then removed from the DNA via alkaline hydrolysis. 75 microliters of 0.25 M EDTA and 0.5 N NaOH were added to each reverse transcription reaction, and the sample was incubated at 95 °C for 10 minutes. The reaction was immediately neutralized by purifying the ssDNA probe with a modified version of the Zymo Oligo Clean and Concentrator protocol. Specifically, the 5-microgram capacity column was replaced with a 25-microgram or 100-microgram capacity DNA column as appropriate. The remainder of the protocol was run according to the manufacturer's instructions. Probe was eluted in 100 microliter RNase-free deionized water and evaporated in a vacuum concentrator. The final pellet was resuspended in 10 microliters of RNase-free water and stored at -20 °C. Denaturing poly-acrylimide gel electrophoresis and absorption spectroscopy revealed that this protocol typically produced 90-100% incorporation of the fluorescent primer into full length probe and 75-90% recovery of the total fluorescent probe. Thus, without exceeding a 150-microliter reaction volume, this protocol can be used to create ~2 nmol of fluorescent probe.

Cell culture and fixation. A549 and IMR90 cells (American Type Culture Collection) were cultured with Dulbecco's Modified Eagle Medium and Eagle's Minimum Essential Medium respectively. Cells were incubated at 37 °C with 5% CO_2 for 36-48 hours. Cells were fixed in 3% paraformaldehyde (Electron Microscopy Sciences) in PBS for 15 minutes, washed with PBS, and permeabilized in 70% ethanol overnight at 4 °C.

Fluorescence In Situ Hybridization (FISH) – primary (encoding) probes. Cells were hydrated in wash buffers (2xSSC, 50% formamide) for 10 minutes, labelled with primary oligos (0.5 nM per sequence) in hybridization buffers (2xSSC, 50% formamide, 1 mg/mL yeast tRNA, and 10% dextran sulfate) overnight at 37 °C, washed with wash buffers at 47 °C for 10 minutes twice, and washed with 2xSSC twice. Fluorescent fiducial beads (Molecular Probes, F-8809) were added at a 1:10,000 dilution in 2xSSC before imaging.

Secondary probes. Secondary (readout) probes (10 nM) were hybridized in secondary hybridization buffers (2xSSC, 20% formamide, and 10% dextran sulfate) to their primary targets for 30 minutes at 37 °C. Cells remained on the microscope stage during the hybridization. An objective heater was used to maintain the temperature at 37 °C. Cells were washed with secondary wash buffers (2xSSC, 20% formamide).

Fluidics and STORM Imaging. Multiple rounds of sequential labeling, washing, imaging, and bleaching were performed on an automated platform consisted of a fluidics setup and a STORM (stochastic optical reconstruction microscopy) microscope. The fluidics setup included a flow chamber (Bioptech FCS2), a peristaltic pump (Rainin Dynamax RP-1), and three computer-controlled 8-way valves (Hamilton MVP and Hamilton HVXM 8-5). This system allowed the automated integration of STORM movie collection and secondary hybridization.

The imaging buffer included, 50 mM Tris (pH 8) 10% (w/v) glucose, 1% βME (2-mercaptoethanol) or 25 mM MEA, with or without 2 mM 1,5-cyclooctadiene, and an oxygen scavenging system (0.5 mg/ml glucose oxidase (Sigma-Aldrich) and 40 microgram/ml catalase (Sigma-Aldrich)). A layer of mineral oil was used to seal the imaging buffer, preventing its acidification over the course of multiple hybridizations.

The STORM setup included an Olympus IX-71 inverted microscope configured for oblique incidence excitation. The samples were continuously illuminated with a 642-nm diode-pumped solid-state laser (VFL-P500-642; MPB communications). A 405-nm solid-state laser (Cube 405-100C; Coherent) was used for activation of dyes. Fluorescence was collected using an Olympus (UPlanSApo 100x, 1.4 NA) objective lens and passed through a custom dichroic, as well as a quad-view beam splitter. All movies were recorded using an EMCCD camera (Andor iAxon 897), imaging at 60Hz. The 512x256 field of view of the camera was split into separate 256x256 pixel movies prior to saving. The left half of this field of view contained the STORM data and the right half contained images of the fluorescent feducial beads. These latter movies were downsampled to 1 Hz prior to saving. During data acquisition, a home-built focus lock was used to maintain a constant focal plane. STORM movies included 20,000 to 30,000 frames in STORM buffer while the bleach movies included 10,000 frames in wash buffer.

Image Analysis – analysis of single-molecule localizations. Movies of single molecule localizations and fluorescent feducial beads were processed separately using a previously published single-emitter localization software.

Image Registration. The starting position of the beads from each round of hybridization were used to align movies from each round. The 2D autocorrelation between bead images of consecutive hybridizations followed by nearest-neighbor matching was used to match beads between images. The pair of beads with the most similar displacement vector were used to compute a rigid translation-rotation warp to align the beads. This alignment method is robust to samples in which multiple feducials are displaced or come detached and reattach during imaging.

Drift Correction. Drift during image acquisition was corrected using the trajectory of the feducial beads (recorded at 1 Hz). Bead positions were linked in each frame. The trajectory of the two beads that moved in the most correlated fashion was taken as the drift trajectory.

mRNA Cluster Calling. Localizations were first screened to be above a threshold number of photons (generally 2000) and required to be within 32 nm of 5 other localizations (parameters may be tuned). The remaining molecule localizations were binned in a 2D histogram of 10x10 nm bins (bin size is a variable parameter). All connected bins were taken to be part of a cluster (diagonal contacts are classified as connected). Clusters were required to have more than 80 total localizations across all hybridizations (variable parameter) to be called an mRNA cluster. The weighted centroids of these clusters from the 2D histogram were recorded as the mRNA positions.

A given cluster is recorded to be represented in an individual hybridization round if more than 9 localizations (variable parameter) are found within a 48 nm radius (variable parameter) of the centroid for that mRNA in each hybridization round.

Cluster Decoding. For each mRNA cluster, a codeword is readout, including "0"s for all the hybridization rounds in which less than the threshold number of localizations are found near the centroid and "1" s for the rounds where above threshold number of localizations are counted. The SECDED codebook decoded these as either perfect matches to target mRNA codewords, correctable errors which can be unambiguously mapped back to target mRNA, or uncorrectable errors, which differed by two or more letters from the words in the codebook.

Fig. 2A shows a STORM image of a cell. Fig. 2B shows a zoom in of the boxed region in Fig. 2A. Each dot indicates a localization. Localizations from different rounds of imaging are shown differently. Fig. 2C shows a representative cluster of localizations from the boxed region in Fig. 2B. The cluster shows localization signals from 4 different hybridizations. This cluster is a putative mRNA encoded with codeword [0 1 0 1 1 1 0 0].

Fig. 2D shows a reconstructed cell image of 14 genes after decoding and error correction. Each gene is shown differently. Fig. 2E shows measured gene expression for the 14 genes from the cell. Fig. 2F shows a comparison of transcript count with ensemble RNA-sequencing data. Fig. 2G shows correlation of transcript expression level between two cells detected using the described approach.

EXAMPLE 4

The following examples are generally directed to multiplexed single-molecule imaging with error-robust encoding allowing for simultaneous measurements of thousands of RNA species in single cells. In general, knowledge of the expression profile and spatial landscape of RNAs in individual cells is essential for understanding the rich repertoire of cellular behaviors. The following examples reports various techniques directed to single-molecule imaging approaches that allow the copy numbers and spatial localizations of thousands of RNA species to be determined in single cells. Some of these techniques are called Multiplexed Error-Robust Fluorescence in Situ Hybridization or "MERFISH."

Using error-robust encoding schemes to combat single-molecule labeling and detection errors, these examples demonstrated the imaging of hundreds to thousands of unique RNA species in hundreds of individual cells. Correlation analysis of the ~10^4 to ~10^6 pairs of genes allowed constraints on gene regulatory networks, prediction of novel functions for many unannotated genes, and identification of distinct spatial distribution patterns of RNAs that correlate with properties of the encoded proteins.

System-wide analyses of the abundance and spatial organization of RNAs in single cells promise to transform understanding in many areas of cell and developmental biology, such as the mechanisms of gene regulation, the heterogeneous behavior of cells, and the development and maintenance of cell fate. Single-molecule fluorescence *in situ* hybridization (smFISH) has emerged as a powerful tool for studying the copy number and spatial organization of RNAs in single cells either in isolation or in their native tissue context. Taking advantage of its ability to map the spatial distributions of specific RNAs with high resolution, smFISH has revealed the importance of subcellular RNA localization in diverse processes such as cell migration, development, and polarization. In parallel, the ability of smFISH to precisely measure the copy numbers of specific RNAs without amplification bias has allowed quantitative measurement of the natural fluctuations in gene expression, which has in turn elucidated the regulatory mechanisms that shape such fluctuations and their role in a variety of biological processes.

However, application of the smFISH approach to many systems-level questions remains limited by the number of RNA species that can be simultaneously measured in single cells. State-of-the-art efforts using combinatorial labeling by either color-based barcodes or sequential hybridization have enabled simultaneous measurements of 10-30 different RNA species in individual cells, yet many interesting biological questions would benefit from the measurement of hundreds to thousands of RNAs within a single cell, which are not achievable using such techniques. For example, analysis of how the expression profile of such a large number of RNAs vary from cell to cell and how these variations correlate among different genes could be used to systematically identify co-regulated genes and map regulatory networks; knowledge of the subcellular organizations of numerous RNAs and their correlations could help elucidate molecular mechanisms underlying the establishment and maintenance of many local cellular structures; and RNA profiling of individual cells in native tissues could allow *in situ* identification of cell type.

The following examples generally discuss certain techniques called MERFISH, which are highly multiplexed smFISH imaging methods that substantially increase the number of RNA species that can be simultaneously imaged in single cells by using combinatorial labeling and sequential imaging with error-robust encoding schemes. These examples demonstrate this multiplexed imaging approach by simultaneously measuring 140 RNA species using an encoding scheme that can both detect and correct errors and 1001 RNA species using an encoding scheme that can detect but not correct errors. It should be understood that these numbers are by way of exemplification only, not limitation. Correlation analyses of the copy number variations and spatial distributions of these genes allowed us to identify groups of genes that are co-regulated and groups of genes that share similar spatial distribution patterns inside the cell.

Combinatorial labeling with error-robust encoding schemes. Combinatorial labeling that identifies each RNA species by multiple (N) distinct signals offers a route to rapidly increase the number of RNA species that can be probed simultaneously in individual cells (Fig. 5A). However, this approach to scaling up the throughput of smFISH to the systems scale faces a significant challenge because not only does the number of addressable RNA species increases exponentially with N, but the detection error rates also grow exponentially with N (Figs. 5B-5D). Imagine a conceptually simple scheme to implement combinatorial labeling, where each RNA species is encoded with a N-bit binary word and the sample is probed with N corresponding rounds of hybridization, each round targeting only the subset of RNAs that should read '1' in the corresponding bit (Fig. 11). N rounds of hybridization

would allow 2^N-1 RNA species to be probed. With just 16 hybridizations, over 64,000 RNA species, which should cover the entire human transcriptome including both messenger RNAs (mRNAs) and non-coding RNAs, could be identified (Fig. 5B; upper symbols). However, as *N* increases, the fraction of RNAs properly detected (the calling rate) would rapidly decrease and, more troublingly, the fraction of RNAs that are identified as incorrect species (the misidentification rate) would rapidly increase (Fig. 5C, lower symbols; Fig. 5D, upper symbols). With realistic error rates per hybridization (measured below), the majority of RNA molecules would be misidentified after 16 rounds of hybridizations!

To address this challenge, error-robust encoding schemes were designed, in which only a subset of the 2^N-1 words separated by a certain Hamming distance were used to encode RNAs. In a codebook where the minimum Hamming distance is 4 (HD4 code), at least four bits must be read incorrectly to change one code word into another (Fig. 12A). As a result, every single-bit error produces a word that is uniquely close to a single code word, allowing such errors to be detected and corrected (Fig. 12B). Double-bit errors produce words with an equal Hamming distance of 2 from multiple code words and, thus, can be detected but not corrected (Fig. 12C). Such a code should substantially increase the calling rate and reduce the misidentification rate (Fig. 5C and 5D, middle symbols). To further account for the fact that it is more likely to miss a hybridization event (an 1-->0 error) than to misidentify a background spot as an RNA (an 0-->1 error) in smFISH measurements, a modified HD4 (MHD4) code was designed, in which the number of '1' bits were kept both constant and relatively low, only four per word, to reduce error and avoid biased detection. This MHD4 code should further increase the calling rate and reduce the misidentification rate (Fig. 5C, upper symbols; Fig. 5D, lower symbols).

In addition to the error considerations, several practical challenges have also made it difficult to probe a large number of RNA species, such as the high cost of the massive number of fluorescently labeled FISH probes needed and the long time required to complete many rounds of hybridization. To overcome these challenges, in this example, a two-step labeling scheme was designed to encode and readout cellular RNAs (Fig. 5E). First, cellular RNAs were labeled with a set of encoding probes (also called primary probes), each probe comprising a RNA targeting sequence and two flanking readout sequences. Four of the *N* unique readout sequences were assigned to each RNA species based on the MHD4 code word of the RNA. Second, these *N* readout sequences were identified with complementary FISH probes, the readout probes (also called secondary probes) via *N* rounds of hybridization and imaging, each round using a unique readout probe. To increase the signal to background

ratio, every cellular RNA was labeled with ~192 encoding probes. Because each encoding probe contained two of the four readout sequences associated with that RNA (Fig. 5E), a maximum of ~96 readout probes can bind to each cellular RNA per hybridization round. To generate the massive number of encoding probes required, they were amplified from array-derived oligonucleotide pools containing tens of thousands of custom sequences using an enzymatic amplification process comprising *in vitro* transcription followed by reverse transcription (Fig. 13, see below regarding probe synthesis). This two-step labeling approach significantly diminished the total hybridization time for an experiment: it was found that efficient hybridization to the readout sequences took only 15 minutes whereas efficient direct hybridization to cellular RNA required more than 10 hours.

Fig. 5 describes MERFISH, a highly multiplexed smFISH approach using combinatorial labeling and error-robust encoding. Fig. 5A shows a schematic depiction of the identification of multiple RNA species in N rounds of imaging. Each RNA species is encoded with a N-bit binary word and during each round of imaging, only the subset of RNAs that should read '1' in the corresponding bit emit signal. Figs. 5B-5D show the number of addressable RNA species (Fig. 5B), the rate at which these RNAs are properly identified (calling rate) (Fig. 5C), and the rate at which RNAs are incorrectly identified as a different RNA species (misidentification rate) (Fig. 5D) plotted as a function of the number of bits (N) in the binary words encoding RNA. In Figs. 5B and 5D, the upper dots are a simple binary code that includes all 2^N-1 possible binary words; the middle dots are the HD4 code where the Hamming distance separating words is 4; and the lower dots are the modified HD4 (MHD4) code where the number of '1' bits are kept at four. These are reversed in Fig. 5C.

The calling and misidentification rates are calculated with per bit error rates of 10% for the 1-->0 error and 4% for the 0-->1 error. Fig. 5E is as schematic diagram of the implementation of a MHD4 code for RNA identification. Each RNA species is first labeled with ~192 encoding probes that convert the RNA into a unique combination of readout sequences (Encoding hyb). These encoding probes each contain a central RNA targeting region flanked by two readout sequences, drawn from a pool of N different sequences, each associated with a specific hybridization round. Encoding probes for a specific RNA species contain a unique combination of four of the N readout sequences, which correspond to the four hybridization rounds where this RNA should read '1'. N subsequent rounds of hybridization with the fluorescent readout probes were used to probe the readout sequences (hyb 1, hyb 2, ..., hyb N). The bound probes were inactivated by photobleaching between

successive rounds of hybridization. For clarity, only one possible pairing of the readout sequences is depicted here for the encoding probes; however, all possible pairs of the four readout sequences are used at the same frequency and distributed randomly along each cellular RNA in the actual experiments.

Fig. 11 shows a schematic description of a combinatorial labeling approach based on a simple binary code. In a conceptually simple labeling approach, 2^N-*1* different RNA species can be uniquely encoded with all *N*-bit binary words (excluding the word with all '0's). In each hybridization round, FISH probes that are targeted to all RNA species that have a '1' in the corresponding bit are included. To increase the ability to discriminate RNA spots from background, each RNA is addressed with multiple FISH probes per hybridization round. Signal from the bound probes is extinguished before the next round of hybridization. This process continues for all *N* hybridization rounds (hyb 1, hyb 2, ...), and all 2^N-*1* RNA species can be identified by the unique on-off pattern of fluorescence signals in each hybridization round.

Fig. 12 shows schematic descriptions of Hamming distance and its use in the identification and correction of errors. Fig. 12A is a schematic representation of a Hamming distance of 4. Figs. 12B and 12C are schematics showing the ability of an encoding scheme with Hamming distance 4 to correct single-bit errors (Fig. 12B) or detect but not correct double-bit errors (Fig. 12C). Arrows highlight bits at which the indicated words differ. Two code words are separated by a Hamming distance of 4 if one of the words has to flip four bits from '1' to '0' or '0' to '1' to convert into the other word. Single-bit error correction is possible because if a measured word differs from a legitimate code word by only one bit, it is most likely an error that arises from misreading this code word, since the code words of all the other RNA species will differ from the measured word by at least three bits. In this case, the measured word can be corrected to a code word that differs by only one bit. If a measured word differs from a legitimate code word by two bits, this measured word can still be identified as an error, but correction is no longer possible since more than one legitimate code word differs from this measured word by two bits.

Fig. 13 shows production of the library of encoding probes. An array-synthesized complex oligopool, containing ~100k sequences, is used as a template for the enzymatic amplification of the encoding probes for different experiments. Each template sequence in the oligopool contains a central target region that can bind to a cellular RNA, two flanking readout sequences, and two flanking index primers. In the first step, the required template molecules for a specific experiment are selected and amplified with an indexed PCR reaction.

To allow amplification via *in vitro* transcription, a T7 promotor is added to the PCR products during this step. In the second step, RNA is amplified from these template molecules via *in vitro* transcription. In the third step, this RNA is reverse transcribed back into DNA. In the final step, the template RNA is removed via alkaline hydrolysis, leaving only the desired ssDNA probes. This protocol produces ~2 nmol of complex pools of encoding probes containing ~20,000 different sequences for the 140-gene experiments or ~100,000 different sequences for the 1001-gene experiments.

EXAMPLE 5

This example illustrates the measurement of 140 genes with MERFISH using a 16-bit MHD4 Code. To test the feasibility of this error-robust, multiplexed imaging approach, this example uses a 140-gene measurement on human fibroblast cells (IMR90) using a 16-bit MHD4 code to encode 130 RNA species while leaving 10 code words as misidentification controls (Figs. 20A to 20H). After each round of hybridization with the fluorescent readout probes, cells were imaged by conventional wide-field imaging with an oblique-incidence illumination geometry. Fluorescent spots corresponding to individual RNAs were clearly detected and were then efficiently extinguished via a brief photobleaching step (Fig. 6A). The sample was stable throughout the 16 rounds of iterative labeling and imaging. The change in the number of fluorescent spots from round to round matched the predicted change based on the relative abundances of RNA species targeted in each round derived from bulk sequencing, and a systematic decreasing trend with increasing number of hybridization rounds was not observed (Fig. 14A). The average brightness of the spots varied from round to round with a standard deviation of 40%, likely due to different binding efficiencies of the readout probes to the different readout sequences on the encoding probes (Fig. 14B). Only a small, systematic decreasing trend in the spot brightness with increasing hybridization rounds was observed, which was on average 4% per round (Fig. 14B).

Next, binary words were constructed from the observed fluorescent spots based on their on-off patterns across the 16 hybridization rounds (Fig. 6B-6D). If the word exactly matched one of the 140 MHD4 code words (exact matches) or differed by only one bit (error-correctable matches), it was assigned to the corresponding RNA species (Fig. 6D). Within the single cell depicted in Figs. 6A and 6B, more than 1500 RNA molecules corresponding to 87% of the 130 encoded RNA species were detected after error correction (Fig. 6E). Similar observations were made in ~400 cells from 7 independent experiments. On average, ~4 times as many RNA molecules and ~2 times as many RNA species were detected per cell after error correction as compared with the values obtained before error correction (Fig. 15).

Two types of errors can occur in the copy number measurement of each RNA species: 1) Some molecules of this RNA species are not detected, leading to a drop in calling rate, and 2) some molecules from other RNA species are misidentified as this RNA species. To assess the extent of misidentification, the 10 misidentification control words were utilized, i.e., code words that were not associated with any cellular RNA. Although matches to these control words were observed, they occurred far less frequently than the real RNA-encoding words: 95% of the 130 RNA-encoding words were counted more frequently than the median count for these control words. Moreover, it was typically found that the ratio of the number of exact matches to the number of matches with one-bit errors for a real RNA-encoding word was substantially higher than the same ratios observed for the misidentification controls, as expected (Figs. 16A and 16B). Using this ratio as a measure of the confidence in RNA identification, it was found that 91% of the 130 RNA species had a confidence ratio greater than the maximum confidence ratio observed for the misidentification controls (Fig. 6F), demonstrating a high accuracy of RNA identification. Subsequent analyses were conducted only on these 91% of genes.

To estimate the calling rate, the error-correction ability of the MHD4 code was utilized to determine the 1-->0 error rates (10% on average) and 0-->1 error rates (4% on average) for each hybridization round (Figs. 16C and 16D). Using these error rates, an ~80% calling rate for individual RNA species after error correction was estimated, i.e. ~80% of the fluorescent spots corresponding to a RNA species were decoded correctly (Fig. 16E). It is noted that although the remaining 20% of spots contributed to a loss in detection efficiency, most of them did not cause species misidentification because they were decoded as double-bit error words and discarded.

To test for potential technical bias in these measurements, the same 130 RNAs species were probed with a different MHD4 codebook by shuffling the code words among different RNA species (Figs. 20A to 20H) and changing the encoding probe sequences. Measurements with this alternative code gave similar misidentification and calling rates (Fig. 17). The copy numbers of individual RNA species per cell measured with these two codebooks showed excellent agreement with a Pearson correlation coefficient of 0.94 (Fig. 6G), indicating that the choice of encoding scheme did not bias the measured counts.

In order to validate the copy numbers derived from the MERFISH experiments, conventional smFISH measurements were performed on 15 of the 130 genes, selected from the full measured abundance range of three orders of magnitude. For each of these genes, both the average copy number and the copy number distribution across many cells agreed

quantitatively between the MERFISH and conventional smFISH measurements (Fig. 18A and 18B). The ratio of the copy numbers determined by these two approaches was 0.82 +/- 0.06 (mean +/- SEM across the 15 measured RNA species, Fig. 18B), which agreed with the estimated 80% calling rate for the multiplexed imaging approach. The quantitative match between this ratio and the estimated calling rate over the full measured abundance range additionally supports the assessment that the misidentification error was low. Given that the agreement between the MERFISH and conventional smFISH results extended to the genes at the lowest measured abundance (<1 copy per cell, Fig. 18B), it was estimated that the measurement sensitivity was at least 1 copy per cell.

As a final validation, the abundance of each RNA species averaged over hundreds of cells was compared to those obtained from a bulk RNA sequencing measurement that were performed on the same cell line. The imaging results correlated remarkably well with bulk sequencing results with a Pearson correlation coefficient of 0.89 (Fig. 6H).

Fig. 6 shows simultaneous measurement of 140 RNA species in single cells using MERFISH with a 16-bit MHD4 code. Fig. 6A shows images of RNA molecules in an IMR90 cell after each hybridization round (hyb 1 – hyb 16). The image after photobleaching (bleach 1) demonstrated efficient removal of fluorescent signals between hybridizations. Fig. 6B shows the localizations of all detected single molecules in this cell colored based on their measured binary words. Inset: the composite fluorescent image of the 16 hybridization rounds for the boxed sub-region with numbered circles indicating potential RNA molecules. A circle indicates an unidentifiable molecule, the binary word of which does not match any of the 16-bit MHD4 code words even after error correction. Fig. 6C shows fluorescent images from each round of hybridization for the boxed sub-region in Fig. 6B with circles indicating potential RNA molecules. Fig. 6D shows corresponding words for the spots identified in Fig. 6C. Crosses represent the corrected bits. Fig. 6E shows the RNA copy number for each gene observed without (lower) or with (higher) error correction in this cell. Fig. 6F shows the confidence ratio measured for the 130 RNA species (left) and the 10 misidentification control words (right) normalized to the maximum value observed from the misidentification controls (dashed line). Fig. 6G is a scatter plot of the average copy number of each RNA species per cell measured with two shuffled codebooks of the MHD4 code. The Pearson correlation coefficient is 0.94 with a p-value of 1×10^{-53}. The dashed line corresponds to the $y = x$ line. Fig. 6H is a scatter plot of the average copy number of each RNA species per cell versus the abundance determined by bulk sequencing in fragments per kilobase per million reads

(FPKM). The Pearson correlation coefficient between the logarithmic abundances of the two measurements was 0.89 with a p-value of $3x10^{-39}$.

Fig. 14 shows the number and average brightness of the fluorescent spots detected in the 16 rounds of hybridization before and after photobleaching. Fig. 14A shows the number of fluorescent spots observed per cell before (higher) and after (lower) photobleaching as a function of hybridization round averaged across all measurements with the first 16-bit MHD4 code. Photobleaching reduces the number of fluorescent spots by two or more orders of magnitude. Hybridization rounds without lower bars represent rounds in which no molecules were observed after bleaching. Also depicted is the expected change in the number of fluorescent spots from round to round (circles) predicted based on the relative abundances of the RNA species targeted in each hybridization round derived from bulk RNA sequencing. The average discrepancy between the observed and predicted number of spots for each hybridization is only 15% of the mean number of spots. This discrepancy does not systematically increase with the number of hybridization rounds. Fig. 14B shows the average brightness of the identified fluorescent spots in each hybridization round averaged across all measurements with the first 16-bit MHD4 code both before (upper) and after (lower) photobleaching. Brightness varies by 40% (standard deviation) across different hybridization rounds. The variation pattern is reproducible between experiments with the same code, likely due to differences in the binding efficiency of the readout probes to the different readout sequences. There is a small systematic trend of decrease in the brightness with increasing hybridization rounds, which is on average 4% per round. Photobleaching extinguishes fluorescence to a level similar to that of the autofluorescence of the cell.

Fig. 15 shows error correction substantially increases the numbers of RNA molecules and RNA species detected in individual cells. Fig. 15A shows a histogram of the ratio of the total number of molecules detected per cell with error correction to the number measured without error correction. Fig. 15B is a histogram of the total number of RNA species detected in each cell with error correction to that without error correction. Both ratios are determined for ~200 cells and the histograms are constructed from these ratios.

Fig. 16 shows characterization of the misidentification and calling rates of RNA species for the 140-gene experiments using a specific 16-bit MHD4 code. Fig. 16A shows the number of measured words exactly matching the code word corresponding to FLNA, represented by the bar in the center of the circle, and the number of measured words with one-bit error compared to the code word of FLNA, represented by the 16 bars on the circle. Fig. 16B is the same as Fig. 16A, but for a code word that was not assigned to any RNA, i.e.,

a misidentification control word. The solid lines connect the exact match to one-bit error words that are generated by 1-->0 errors. Based on the observation that the ratio of the number of exact matches to the number of error-correctable matches for a real RNA-encoding word was typically substantially higher than the same ratios observed for the misidentification controls, this ratio was defined as a confidence ratio for RNA identification. The confidence ratio measured for all 130 RNA species (center bar) and 10 misidentification control words not assigned to any RNA (outer bars) using this 16-bit MHD4 code is show in Fig. 6F. Fig. 16C and 16D show the average error rates for the 1-->0 error (Fig. 16C) and 0-->1 error (Fig. 16D) for each hybridization round. Fig. 16E shows the calling rate for each RNA species estimated from the 1-->0 and 0-->1 error rates. Genes are sorted from left to right based on the measured abundance, which spans three orders of magnitude. The calling rates are largely independent of the abundance of the gene.

Fig. 17 shows characterization of the misidentification and calling rates for a second 16-bit MHD4 code. In this second encoding scheme, the 140 code words were shuffled among different RNA species and changed the encoding probe sequences. Fig. 17A shows the normalized confidence ratio measured for the 130 RNA species (left) and the 10 misidentification control words not assigned to any RNA (right). The normalized confidence ratio is determined the same way as in Fig. 6F. Figs. 17B and 17C show the average error rates determined for the 1-->0 error (Fig. 17B) and 0-->1 error (Fig. 17C) for each hybridization round. Fig. 17D shows the calling rate determined for each RNA species estimated from the 1-->0 and 0-->1 error rates. Genes are sorted from left to right based on the measured abundance.

Fig. 18 shows a comparison of the MERFISH measurements with conventional smFISH results for a subset of genes. Fig. 18A shows the distributions of RNA copy numbers in single cells for three example genes KIAA1199, DYNC1H1, and LMTK2 in the high, medium, and low abundance ranges, respectively. Lighter bars: distributions constructed from ~400 cells in the 140-gene measurements using the MHD4 codes. Darker bars: distributions constructed from ~100 cells in the conventional smFISH measurements. Fig. 18B shows a comparison of the average RNA copy numbers per cell measured in the 140-gene experiments using the MHD4 codes to those determined by conventional smFISH for 15 genes. The average ratio of the copy number measured using the MHD4 measurements to that measured using conventional smFISH was 0.82 +/- 0.06 (mean +/- SEM across 15 genes). The dashed line corresponds to the $y = x$ line.

Figs. 20A to 20H show two different codebooks for the 140-gene experiments. The specific code words of the 16-bit MHD4 code assigned to each RNA species studied in the two shuffles of the 140-gene experiment. The "Genes" columns contain the name of the gene. The "Codewords" columns contain the specific binary word assigned to each gene.

EXAMPLE 6

This example is generally directed to high-throughput analysis of cell-to-cell variation in gene expression. The MERFISH approach allows parallelization of measurements of many individual RNA species and co-variation analysis between different RNA species. In this example, the parallelization aspect was first illustrated by examining the cell-to-cell variation in the expression level of each of the measured genes (Fig. 7A). To quantify the measured variation, the Fano factors, defined as the ratio of the variance to the mean RNA copy number, were computed for all measured RNA species. The Fano factors substantially deviated from 1, the value expected for a simple Poisson process, for many genes and exhibited an increasing trend with the mean RNA abundance (Fig. 7B). This trend of increasing Fano factors with mean RNA abundance can be explained by changes in the transcription rate and/or promoter off-switching rates but not by changes in the promoter on-switching rate.

Moreover, several RNA species were identified with substantially larger Fano factors than this average trend. For example, it was found that SLC5A3, CENPF, MKI67, TNC and KIAA1199 displayed Fano factor values substantially higher than those of the other genes expressed at similar abundance levels. The high variability of some of these genes can be explained by their association with the cell cycle. For example, two of these particularly 'noisy' genes MKI67 and CENPF were both annotated as cell-cycle related genes, and based on their bimodal expression (Fig. 7C), it is proposed that their transcription is strongly regulated by the cell cycle. Other high-variability genes did not show the same bimodal expression patterns and are not known to be associated with the cell cycle.

Analysis of co-variations in the expression levels of different genes can reveal which genes are co-regulated and elucidate gene regulatory pathways. At the population level, such analysis often requires the application of external stimuli to drive gene expression variation; hence, correlated expression changes can be observed among genes that share common regulatory elements influenced by the stimuli. At the single-cell level, one can take advantage of the natural stochastic fluctuations in gene expression for such analysis and can thus study multiple regulatory networks without having to stimulate each of them

individually. Such co-variation analysis can constrain regulatory networks, suggest new regulatory pathways, and predict function for unannotated genes based on associations with co-varying genes.

This approach was applied to the 140-gene measurements and the ~10,000 pairwise correlation coefficients describing how the expression levels of each pair of genes co-varied from cell to cell were examined. Many of the highly variable genes showed tightly correlated or anti-correlated variations (Fig. 7C). To better understand the correlations for all gene pairs, a hierarchical clustering approach was adopted to organize these genes based on their correlation coefficients (Fig. 7D). From the cluster tree structure, seven groups of genes with substantially correlated expression patterns were identified (Fig. 7D). Within each of the seven groups, every gene showed significantly stronger average correlation with other members of the group than with genes outside the group. To further validate and understand these groups, gene ontology (GO) terms enriched in each of these seven groups were identified. Notably, the enriched GO terms within each group shared similar functions and were largely unique to each group (Fig. 7E), validating the notion that the observed co-variation in expression reflects some commonalities in the regulation of these genes.

This example describes two of these groups as illustrative examples. The predominant GO terms associated with Group 1 were terms associated with the extracellular matrix (ECM) (Figs. 7D and 7E). Notable members of this group included ECM components, such as FBN1, FBN2, COL5A, COL7A and TNC, and glycoproteins linking the ECM and cell membranes, such as VCAN and THBS1. The group also included an unannotated gene, KIAA1199, which may predict to play a role in ECM metabolism based on its association with this cluster. Indeed, this gene has recently been identified as an enzyme involved in the regulation of hyaluronan, a major sugar component of the ECM.

Group 6 contained many genes that encode vesicle transport proteins and proteins associated with cell motility (Figs. 7D and 7E). The vesicle transport genes included microtubule motors and related genes DYNC1H, CKAP1, and factors associated with vesicle formation and trafficking, like DNAJC13 and RAB3B. Again, an unannotated gene, KIAA1462, was found within this cluster. Based on its strong correlation with DYNC1H1 and DNAJC13, it is predicted that this gene may be involved in vesicle transport. The cell motility genes in this group included actin-binding proteins like AFAP1, SPTAN1, SPTBN1, and MYH10, and genes involved in the formation of adhesion complexes, like FLNA and FLNC. Several GTPase-associated factors involved in the regulation of cell motility, attachment and contraction also fell into this group, including DOCK7, ROCK2, IQGAP1,

PRKCA, and AMOTL1. The observation that some cell motility genes correlated with vesicle transport genes is consistent with the role of vesicle transport in cell migration. An additional interesting feature of group 6 is that a subset of these genes, in particular those related to cell motility, were anti-correlated with members of the ECM group discussed above (Fig. 7D). This anti-correlation may reflect regulatory interactions that mediate switching of cells between adherent and migratory states.

Fig. 7 shows cell-to-cell variations and pairwise correlations for the RNA species determined from the 140 gene measurements. Fig. 7A shows a comparison of gene expression levels in two individual cells. Fig. 7B shows Fano factors for individual genes. Error bars represent standard error of the mean determined from 7 independent data sets. Fig. 7C shows Z-scores of the expression variations of four example pairs of genes showing correlated (top two) or anti-correlated (bottom two) variation for 100 randomly selected cells. The Z-score is defined as the difference from the mean normalized by the standard deviation. Fig. 7D is a matrix of the pairwise correlation coefficients of the cell-to-cell variation in expression for the measured genes, shown together with the hierarchical clustering tree. The seven groups identified by a specific threshold on the cluster tree (dashed line) are indicated by the black boxes in the matrix and lines on the tree, with grey lines on the tree indicating ungrouped genes. Different threshold choices on the cluster tree could be made to select either smaller subgroups with tighter correlations or larger super-groups containing more weakly coupled subgroups. Two of the seven groups are enlarged on the right. Fig. 7E shows enrichment of 30 selected, statistically significantly enriched GO terms in the seven groups. Enrichment refers to the ratio of the fraction of genes within a group that have the specific GO term to the fraction of all measured genes having that term. Not all of the GO terms presented here are in the top 10 list.

EXAMPLE 7

This example illustrates mapping spatial distributions of RNAs. As an imaging based approach, MERFISH also allowed the investigation of the spatial distributions of many RNA species simultaneously. Several patterns emerged from the visual inspection of individual genes, with some RNA transcripts enriched in the perinuclear region, some enriched in the cell periphery, and some scattered throughout the cell (Fig. 8A). To identify genes with similar spatial distributions, the correlation coefficients for the spatial density profiles of all pairs of RNA species were determined, and these RNAs were organized based on the pairwise correlations again using the hierarchical clustering approach. The correlation coefficient matrix showed groups of genes with correlated spatial organizations, and the two

most notable groups with the strongest correlations are indicated in Fig. 8B. Group I RNAs appeared enriched in the perinuclear region whereas group II RNAs appeared enriched near the cell peripheral region (Fig. 8C). Quantitative analysis of the distances between each RNA molecule and the cell nucleus or the cell periphery indeed confirmed this visual impression (Fig. 8D).

Group I contained genes encoding extracellular proteins such as FBN1, FBN2 and THSB1, secreted proteins such as PAPPA, and integral membrane proteins such as LRP1 and GPR107. These proteins have no obvious commonalities in function. Rather a GO analysis showed significant enrichment for location terms, such as extracellular region, basement membrane, or perivitelline space (Fig. 8E). To reach these locations, proteins must pass through the secretion pathway, which often requires translation of mRNA at the endoplasmic reticulum (ER). Thus, it is proposed that the spatial pattern that were observed for these mRNAs reflects their co-translational enrichment at the ER. The enrichment of these mRNAs in the perinuclear region (Figs. 8C and 8D, lighter shading), where the rough ER resides, supports this conclusion.

Group II contained genes encoding the actin-binding proteins, including filamins FLNA and FLNC, talin TLN1, and spectrins SPTAN1 and SPTBN1; the microtubule-binding protein CKAP5; and the motor proteins MYH10 and DYNC1H1. This group was enriched with GO terms such as cortical actin cytoskeleton, actin filament binding, and cell-cell adherens junction (Fig. 8E). Beta-actin mRNA may be enriched near the cell periphery in fibroblasts as are mRNAs that encode members of the actin-binding Arp2/3 complex. The enrichment of group II mRNAs in the peripheral region of the cells (Figs. 8C and 8D) suggests that the spatial distribution of the Group II genes might be related to the distribution of actin cytoskeleton mRNAs.

Fig. 8 shows distinct spatial distributions of RNAs observed in the 140-gene measurements. Fig. 8A shows examples of the spatial distributions observed for four different RNA species in a cell. Fig. 8B is a matrix of the pairwise correlation coefficients describing the degree with which the spatial distributions of each gene pair is correlated, shown together with the hierarchical clustering tree. Two strongly correlating groups are indicated by the black boxes on the matrix and shading on the tree. Fig. 8C shows the spatial distributions of all RNAs in the two groups in two example cells. Lighter symbols: group I genes; darker symbols: group II genes. Fig. 8D shows average distances for genes in group I and genes in group II to the cell edge or the nucleus normalized to the average distances for

all genes. Error bars represent SEM across 7 data sets. Fig. 8E shows enrichment of GO terms in each of the two groups.

EXAMPLE 8

This example illustrates measuring 1001 genes with a 14-bit MHD2 code. This example further increases the throughput of MERFISH measurements by simultaneously imaging ~1000 RNA species. This increase could be achieved with the MHD4 code by increasing the number of bits per code word to 32 while maintaining the number of '1' bits per word at four (Fig. 5B). While the stability of the samples across many hybridization rounds (Fig. 14) suggests that such an extension is potentially feasible, an alternative approach is shown here that did not require an increase in the number of hybridizations by relaxing the error correction requirement but keeping the error detection capability. For example, by reducing the Hamming distance from 4 to 2, all 14-bit words could be used that contain four '1' bits to encode 1001 genes and these RNAs were probed with only 14 rounds of hybridization. However, because a single error can produce a word equally close to two different code words, error correction is no longer possible for this modified Hamming-distance-2 (MHD2) code. Hence, it was expected that the calling rate would be lower and the misidentification rate to be higher with this encoding scheme.

To evaluate the performance of this 14-bit MHD2 code, 16 of the 1001 possible code words were set aside as misidentification controls and used the remaining 985 words to encode cellular RNAs. Among these 985 RNAs included 107 RNA species probed in the 140-gene experiments as an additional control. The 1001-gene experiments were performed in IMR90 cells using a similar procedure as described above. To allow all encoding probes to be synthesized from a single 100,000-member oligopool, the number of encoding probes per RNA species was reduced to ~94. Fluorescent spots corresponding to individual RNA molecules were again detected in each round of hybridization with the readout probes and, based on their on-off patterns, these spots were decoded into RNA (Figs. 9A, 19A and 19B). 430 RNA species were detected in the cell shown in Fig. 9A, and similar results were obtained in ~200 imaged cells in 3 independent experiments.

As expected, the misidentification rate of this scheme was higher than that of the MHD4 code. 77% of all real RNA words were detected more frequently than the median count for the misidentification controls instead of the 95% value observed in the MHD4 measurements. Using the same confidence ratio analysis as described above, it was found that 73% (instead of 91% for the MHD4 measurements) of the 985 RNA species were measured with a confidence ratio larger than the maximum value observed for the

misidentification controls (Fig. 19C). RNA copy numbers measured from these 73% RNA species showed excellent correlation with the bulk RNA sequencing results (Pearson correlation coefficient r = 0.76; Fig. 9B, black). It is worth noting that the remaining 27% of the genes still exhibit good, albeit lower, correlation with the bulk RNA sequencing data (r = 0.65; Fig. 9B, red), but the conservative measure of excluding them from further analysis was taken.

The lack of an error correction capability also decreased the calling rate of each RNA species: When comparing the 107 RNA species common in both the 1001-gene and 140-gene measurements, it was found that the copy numbers per cell of these RNA species were lower in the 1001-gene measurements (Figs. 9C and 19D). The total count of these RNAs per cell was ~1/3 of that observed in the 140-gene measurements. Thus the lack of error correction in the MHD2 code produced a ~3-fold decrease in the calling rate, which is consistent with the ~4-fold decrease in calling rate observed for the MHD4 code when error correction was not applied. As expected from the quantitative agreement between 140-gene measurements and conventional smFISH results, comparison of the 1001-gene measurements with conventional smFISH results for 10 RNA species also indicated a ~3-fold drop in calling rate (Fig. 18C). Despite the expected reduction in calling rate, the good correlations found between the copy numbers observed in the 1001-gene measurements and those observed in the 140-gene measurements, as well as in conventional smFISH and bulk RNA sequencing measurements, indicates that the relative abundance of these RNAs can be quantified with the MHD2 encoding scheme.

Simultaneously imaging ~1000 genes in individual cells substantially expanded the ability to detect co-regulated genes. Fig. 10A shows the matrix of pairwise correlation coefficients determined from the cell-to-cell variations in the expression levels of these genes. Using the same hierarchical clustering analysis as described above, ~100 groups of genes with correlated expression were identified. Remarkably, nearly all of these ~100 groups showed statistically significant enrichment of functionally related GO terms (Fig. 10B). These included some of the groups identified in the 140-gene measurements, such as the group associated with cell replication genes and the group associated with cell motility genes (Figs. 10A and 10B, groups 7 and 102), as well as many new groups. The groups identified here included 46 RNA species lacking any previous GO annotations, for which function based on their group association may be hypothesized. For example, KIAA1462 is part of the cell motility group, as also shown in the 140-gene experiments, suggesting a potential role of this gene in cell motility (Fig. 10A, group 102). Likewise, KIAA0355 is part

of a new group enriched in genes associated with heart development (Fig. 10A, group 79), and C17orf70 is part of a group associated with ribosomal RNA processing (Fig. 10A, group 22). Using these groupings, cellular functions for 61 transcription factors and other partially annotated proteins of unknown functions may be hypothesized. For example, the transcription factors Z3CH13 and CHD8 are both members of the cell motility group, suggesting their potential role in the transcriptional regulation of cell motility genes.

Fig. 9 shows simultaneous measurements of 1001 genes in single cells using MERFISH with a 14-bit MHD2 code. Fig. 9A shows the localizations of all detected single molecules in a cell colored based on their measured binary words. Inset: the composite, false-colored fluorescent image of the 14 hybridization rounds for the boxed sub-region with numbered circles indicating potential RNA molecules. Circles indicate unidentifiable molecules, the binary words of which do not match any of the 14-bit MHD2 code words. Images of individual hybridization round are shown in Fig. 19A. Fig. 9B is a scatter plot of the average copy number per cell measured in the 1001-gene experiments versus the abundance measured via bulk sequencing. The upper symbols are for the 73% of genes detected with confidence ratios higher than the maximum ratio observed for the misidentification controls. The Pearson correlation coefficient is 0.76 with a p-value of $3x10^{-133}$. The lower symbols are for the remaining 27% of genes. The Pearson correlation coefficient is 0.65 with a p-value of $3x10^{-33}$. Fig. 9C is a scatter plot of the average copy number for the 107 genes shared in both the 1001-gene measurement with the MHD2 code and the 140-gene measurement with the MHD4 code. The Pearson correlation coefficient is 0.89 with a p-value of $9x10^{-30}$. The dashed line corresponds to the $y = x$ line.

Fig. 10 shows co-variation analysis of the RNA species measured in the 1001-gene measurements. Fig. 10A is a matrix of all pairwise correlation coefficients of the cell-to-cell variation in expression for the measured genes shown with the hierarchical clustering tree. The ~100 identified groups of correlated genes are indicated by shading on the tree. Zoom in of four of the groups described in the text are shown on the right. Fig. 10B is an enrichment of 20 selected, statistically significantly enriched GO terms in the four groups.

Fig. 19 shows decoding and error assessment of the 1001-gene experiments. Fig. 19A shows images of the boxed sub-region of the cell in Fig. 9A for each of 14 hybridization rounds. The final panel is a composite image of these 14 rounds. Circles indicate fluorescent spots that have been identified as potential RNA molecules. Some circles in the composite image indicate unidentifiable molecules, the binary words of which do not match any of the 14-bit MHD2 code words. Fig. 9B shows the corresponding binary word for each of the

spots identified in Fig. 9A with the RNA species to which it is decoded. 'unidentified' implies that the measured binary word does not match any of the 1001 code words. Fig. 19C shows the normalized confidence ratios measured for the 985 RNA species (left) and the 16 misidentification control words not targeted to any RNA (right). The normalized confidence ratio is defined as in Fig. 6F. Fig. 19D shows a histogram of the reduction in detected abundance of 107 genes present in both the 1001-gene experiments and the 140-gene experiments. "Fold decrease in copy number" is defined as the average number of RNA molecules per cell for each species measured in the 140-gene experiments divided by the corresponding average number measured in the 1001-gene experiments.

Fig. 18C is a comparison of the average RNA copy numbers per cell measured in the 1001-gene experiments using the MHD2 code to those determined by conventional smFISH for 10 genes. The average ratio of the copy number measured using the MHD2 measurements to that measured using conventional smFISH was 0.30 +/- 0.05 (mean +/- SEM across 10 genes). The dashed line corresponds to the $y = x$ line and the dotted line corresponds to the $y = 0.30x$ line.

EXAMPLE 9

The above examples illustrate a highly multiplexed detection scheme for systems-level RNA imaging in single cells. Using combinatorial labeling, sequential hybridization and imaging, and two different error-robust encoding schemes, either 140 or 1001 genes in hundreds of individual human fibroblast cells were simultaneously imaged. Of the two encoding schemes presented here, the MHD4 code is capable of both error detection and error correction, and hence can provide a higher calling rate and a lower misidentification rate than the MHD2 code, which instead can only detect but cannot correct errors. MHD2, on the other hand, provides a faster scaling of the degree of multiplexing with the number of bits than MHD4. Other error-robust encoding schemes can also be used for such multiplexed imaging, and experimenters can set the balance between detection accuracy and ease of multiplexing based on the specific requirements of the experiments.

By increasing the number of bits in the code words, it should be possible to further increase the number of detectable RNA species using MERFISH with, for example, a MHD4 or MHD2 code. For example, using the MHD4 code with 32 total bits and four or six '1' bits would increase the number of addressable RNA species to 1,240 or 27,776, respectively. The latter is the approximate scale of the human transcriptome. The predicted misidentification and calling rates are still reasonable for the 32-bit MHD4 code (shown in Figs. 5C and 5D for the MHD4 code with four '1' bits and similar rates were calculated for the MHD4 code with

six '1' bits). If more accurate measurements are desired, an additional increase in the number of bits would allow the use of encoding schemes with a Hamming distance greater than 4, further enhancing the error detection and correction capability. While an increase in the number of bits by adding more hybridization rounds would increase the data collection time and potentially lead to sample degradation, these problems could be mitigated by utilizing multiple colors to readout multiple bits in each round of hybridization.

As the degree of multiplexing is increased, it is important to consider the potential increase in the density of RNAs that need to be resolved in each round of imaging. Based on the imaging and sequencing results, it can be estimated that including the whole transcriptome of the IMR90 cells would lead to a total RNA density of ~200 molecules/micrometer^3. Using the current imaging and analysis methods, 2-3 molecules/ micrometer^3 per hybridization round could be resolved, which would reach a total RNA density of ~20 molecules/micrometer^3 after 32 rounds of hybridization. This density should allow all but the top 10% most expressed genes to be imaged simultaneously or a subset of genes with even higher expression levels to be included. By utilizing more advanced image analysis algorithms to better resolve overlapping images of individual molecules, such as compressed sensing, it is possible to extend the resolvable density by ~4-fold and thus allow all but the top 2% most expressed genes to be imaged all together.

These examples have illustrated the utility of the data derived from highly multiplexed RNA imaging by using co-variation and correlation analysis to reveal distinct sub-cellular distribution patterns of RNAs, to constrain gene regulatory networks, and to predict functions for many previously unannotated or partially annotated genes with unknown functions. Given its ability to quantify RNAs across a wide range of abundances without amplification bias while preserving native context, systems and methods such as MERFISH will allow many applications of *in situ* transcriptomic analyses of individual cells in culture or complex tissues.

EXAMPLE 10

Following are various materials and methods used in the above examples.

Probe design. Each RNA species in the target set was randomly assigned a binary code word either from all 140 possible code words of the 16-bit MHD4 code or from all 1001 possible code words of the 14-bit MHD2 code.

Array-synthesized oligopools were used as templates to make the encoding probes. The template molecule for each encoding probe contained three components: i) a central targeting sequence for *in situ* hybridization to the target RNA, ii) two flanking readout

sequences designed to hybridize each of two distinct readout probes, and iii) two flanking primer sequences to allow enzymatic amplification of the probes (Fig. 13). The readout sequences were taken from the 16 possible readout sequences, each corresponding to one hybridization round. The readout sequences were assigned to the encoding probes such that for any RNA species each of the 4 readout sequences were distributed uniformly along the length of the target RNA and appeared at the same frequency. Template molecules for the 140-gene library also included a common 20-nucleotide (nt) priming region between the first PCR primer and the first readout sequence. This priming sequence was used for the reverse transcription step described below.

Multiple experiments were embedded in a single array-synthesized oligopool, and PCR was used to selectively amplify only the oligos required for a specific experiment. Primer sequences for this indexed PCR reaction were generated from a set of orthogonal 25-nt sequences. These sequences were trimmed to 20 nt and selected for i) a narrow melting temperature range (70 °C to 80 °C), ii) the absence of consecutive repeats of 3 or more identical nucleotides, and iii) the presence of a GC clamp, i.e. one of the two 3' terminal bases must be G or C. To further improve specificity, these sequences were then screened against the human transcriptome using BLAST+, and primers with 14 or more contiguous bases of homology were eliminated. Finally, BLAST+ was again used to identify and exclude primers that had an 11-nt homology region at the 3' end of any other primer or a 5-nt homology region at the 3' end of the T7 promoter. The forward primer sequences (Primer 1) were determined as described above, whereas the reverse primers each contain a 20-nt sequence as described above plus a 20-nt T7 promoter sequence to facilitate amplification via *in vitro* transcription (Primer 2). The primer sequences used in the 140-gene and 1001-gene experiments are listed below.

Table 2

Experiment Name	Primer 1 Sequence (Index Primer 1)	Primer 2 Sequence (T7 promoter plus the reverse complement of Index Primer 2)
140-gene Codebook 1	GTTGGTCGGCACTTGGGTGC (SEQ ID NO: 18)	TAATACGACTCACTATAGGGAAAGCCGGTTCATCCGGTGG (SEQ ID NO: 21)
140-gene Codebook 2	CGATGCGCCAATTCCGGTTC (SEQ ID NO: 19)	TAATACGACTCACTATAGGGTGATCATCGCTCGCGGGTTG (SEQ ID NO: 22)
1001-gene	CGCGGGCTATATGCGAACC	TAATACGACTCACTATAGGGCGTGGAGG

G GCATACAACGC (SEQ ID NO: 23)
(SEQ ID NO: 20)

30-nt-long readout sequences were created by concatenating fragments of the same orthogonal primer set generated above by combining one 20-nt primer with a 10-nt fragment of another. These readout sequences were then screened, using BLAST+, for orthogonality with the index primer sequences and other readout sequences (no more than 11 nt of homology) and for potential off-target binding sites in the human genome (no more than 14 nt of homology). Fluorescently labeled readout probes with sequences complementary to the readout sequences were used to probe these readout sequences, one in each hybridization round. All used readout probes sequences are listed below.

Table 3

Bit	Readout probes	
1	CGCAACGCTTGGGACGGTTCCAATCGGATC/3Cy5Sp/	SEQ ID NO: 24
2	CGAATGCTCTGGCCTCGAACGAACGATAGC/3Cy5Sp/	SEQ ID NO: 25
3	ACAAATCCGACCAGATCGGACGATCATGGG/3Cy5Sp/	SEQ ID NO: 26
4	CAAGTATGCAGCGCGATTGACCGTCTCGTT/3Cy5Sp/	SEQ ID NO: 27
5	GCGGGAAGCACGTGGATTAGGGCATCGACC/3Cy5Sp/	SEQ ID NO: 28
6	AAGTCGTACGCCGATGCGCAGCAATTCACT/3Cy5Sp/	SEQ ID NO: 29
7	CGAAACATCGGCCACGGTCCCGTTGAACTT/3Cy5Sp/	SEQ ID NO: 30
8	ACGAATCCACCGTCCAGCGCGTCAAACAGA/3Cy5Sp/	SEQ ID NO: 31
9	CGCGAAATCCCCGTAACGAGCGTCCCTTGC/3Cy5Sp/	SEQ ID NO: 32
10	GCATGAGTTGCCTGGCGTTGCGACGACTAA/3Cy5Sp/	SEQ ID NO: 33
11	CCGTCGTCTCCGGTCCACCGTTGCGCTTAC/3Cy5Sp/	SEQ ID NO: 34
12	GGCCAATGGCCCAGGTCCGTCACGCAATTT/3Cy5Sp/	SEQ ID NO: 35
13	TTGATCGAATCGGAGCGTAGCGGAATCTGC/3Cy5Sp/	SEQ ID NO: 36
14	CGCGCGGATCCGCTTGTCGGGAACGGATAC/3Cy5Sp/	SEQ ID NO: 37
15	GCCTCGATTACGACGGATGTAATTCGGCCG/3Cy5Sp/	SEQ ID NO: 38
16	GCCCGTATTCCCGCTTGCGAGTAGGGCAAT/3Cy5Sp/	SEQ ID NO: 39

The readout probes used for the 140-gene libraries were probes 1 through 16. The readout probes used for the 1001-gene experiment were probes 1 through 14. /3Cy5Sp/ indicates a 3' Cy5 modification.

To design the central targeting sequences of the encoding probes, the abundance of different transcripts in IMR90 cells using Cufflinks v2.1, total RNA data from the ENCODE project, and human genome annotations from Gencode v18 were complied. Probes were designed from gene models corresponding to the most abundant isoform using OligoArray2.1 with the following constraints: the target sequence region is 30-nt long; the melting temperatures of the hybridized region of the probe and cellular RNA target is greater than 70 °C; there is no cross hybridization targets with melting temperatures greater than 72 °C; there

is no predicted internal secondary structures with melting temperatures greater than 76 °C; and there is no contiguous repeats of 6 or more identical nucleotides. Melting temperatures were adjusted to optimize the specificity of these probes and minimize secondary structure while still producing sufficient numbers of probes for the libraries. To decrease computational cost, isoforms were divided into 1-kb regions for probe design. Using BLAST+, all potential probes that mapped to more than one cellular RNA species were rejected. Probes with multiple targets on the same RNA were kept.

For each gene in the 140-gene experiments, 198 putative encoding probe sequences were generated by concatenating the appropriate index primers, readout sequences, and targeting regions as shown in Fig. 13. To address the possibility that concatenation of these sequences introduced new regions of homology to off-target RNAs, BLAST+ was used to screen these putative sequences against all human rRNA and tRNA sequences as well as highly expressed genes (genes with FPKM > 10,000). Probes with greater than 14 nt of homology to rRNAs or tRNAs or greater than 17 nt of homology to highly expressed genes were removed. After these cuts, there were ~192 (with a standard deviation of 2) probes per gene for both MHD4 codebooks used in the 140-gene experiments. The same protocol for the 1001-gene experiments was used, as follows: Starting with 96 putative targeting sequences per gene, ~94 (with a standard deviation of 6) encoding probes per gene were obtained after these additional homology cuts. The number of encoding probes per RNA was decreased for the 1001-gene experiments so that these probes could be synthesized from a single 100,000-member oligopool as opposed to two separate pools. Each encoding probe was designed to contain two of the four readout sequences associated with each code word, hence only half of the bound encoding probes can bind readout probe during any given hybridization round. ~192 or ~94 encoding probes per RNA were used to obtain high signal-to-background ratios for individual RNA molecules. The number of encoding probes per RNA could be substantially reduced but still allow single RNA molecules to be identified. In addition, increasing the number of readout sequences per encoding probe or using optical sectioning methods to reduce the fluorescence background may allow further reduction in the number of the encoding probes per RNA.

Two types of misidentification controls were designed. The first control (blank words) were not represented with encoding probes. The second type of control (no-target words) had encoding probes that were not targeting any cellular RNA. The targeting regions of these probes were composed of random nucleotide sequences subject to the same constraints used to design the RNA targeting sequences described above. Moreover, these

random sequences were screened against the human transcriptome to ensure that they contain no significant homology (>14-nt) to any human RNA. The 140-gene measurements contained 5 blank words and 5 no-target words. The 1001-gene measurements contained 11 blank words and 5 no-target words.

Probe synthesis. The encoding probes were synthesized using the following steps, and this synthesis protocol is illustrated in Fig. 13.

Step 1: The template oligopool (CustomArray) was amplified via limited-cycle PCR on a Bio-Rad CFX96 using primer sequences specific to the desired probe set. To facilitate subsequent amplification via *in vitro* transcription, the reverse primer contained the T7 promoter. All primers were synthesized by IDT. This reaction was column purified (Zymo DNA Clean and Concentrator; D4003).

Step 2: The purified PCR products were then further amplified ~200-fold and converted into RNA via a high yield *in vitro* transcription according to the manufacturer's instructions (New England Biolabs, E2040S). Each 20 microliter reaction contained ~1 microgram of template DNA from above, 10 mM of each NTP, 1x reaction buffer, 1x RNase inhibitor (Promega RNasin, N2611) and 2 microliters of the T7 polymerase. This reaction was incubated at 37 °C for 4 hours to maximize yield. This reaction was not purified before the following steps.

Step 3: The RNA products from the above *in vitro* transcription reaction were then converted back into DNA via a reverse transcription reaction. Each 50 microliter reaction contained the unpurified RNA produce from Step 2 supplemented with 1.6 mM of each dNTP, 2 nmol of a reverse transcription primer, 300 units of Maxima H- reverse transcriptase (Thermo Scientific, EP0751), 60 units of RNasin, and a final 1x concentration of the Maxima RT buffer. This reaction was incubated at 50 °C for 45 minutes, and the reverse transcriptase was inactivated at 85 °C for 5 minutes. The templates for the 140-gene libraries contain a common priming region for this reverse transcription step; thus, a single primer was used for this step when creating these probes. Its sequence was CGGGTTTAGCGCCGGAAATG (SEQ ID NO: 40). A common priming region was not included for the 1001-gene library; thus, the reverse transcription was conducted with the forward primer: CGCGGGCTATATGCGAACCG (SEQ ID NO: 20).

Step 4: To remove the template RNA, 20 microliters of 0.25 M EDTA and 0.5 N NaOH was added to the above reaction to selectively hydrolyze RNA, and the sample was incubated at 95 °C for 10 minutes. This reaction was then immediately purified by column purification using a 100-microgram-capacity column (Zymo Research, D4030) and the Zymo

Oligo Clean and Concentrator protocol. The final probes were eluted in 100 microliters of RNase-free deionized water, evaporated in a vacuum concentrator, and then resuspended in 10 μL of encoding hybridization buffer (see below). Probes were stored at -20 °C. Denaturing poly-acrylamide gel electrophoresis and absorption spectroscopy were used to confirm the quality of the probes and revealed that this probe synthesis protocol converts 90-100% of the reverse-transcription primer into full length probe and of the probe that is constructed, 70-80% is recovered during the purification step.

Fluorescently labeled readout probes have sequences complementary to the readout sequences described above and a Cy5 dye attached at the 3' end. These probes were obtained from IDT and HPLC purified.

Sample preparation and labeling with encoding probes. Human primary fibroblasts (American Type Culture Collection, IMR90) were used in this work. These cells are relatively large and flat, facilitating wide-field imaging without the need for optical sectioning. Cells were cultured with Eagle's Minimum Essential Medium. Cells were plated on 22-mm, #1.5 coverslips (Bioptechs, 0420-0323-2) at 350,000 cells/coverslip and incubated at 37 °C with 5% CO_2 for 48-96 hours within petri dishes. Cells were fixed for 20 minutes in 4% paraformaldehyde (Electron Microscopy Sciences, 15714) in 1x phosphate buffered saline (PBS; Ambion, AM9625) at room temperature, reduced for 5 minutes with 0.1% w/v sodium borohydride (Sigma, 480886) in water to reduce background fluorescence, washed three times with ice-cold 1x PBS, permeabilized for 2 minutes with 0.5% v/v Triton (Sigma, T8787) in 1x PBS at room temperature, and washed three times with ice cold 1x PBS.

Cells were incubated for 5 minutes in encoding wash buffer comprising 2x saline-sodium citrate buffer (SSC) (Ambion, AM9763), 30% v/v formamide (Ambion, AM9342), and 2 mM vanadyl ribonucleoside complex (NEB, S1402S). 10 microliters of 100 micromolar(140-gene experiments) or 200 micromolar (1001-gene experiments) encoding probes in encoding hybridization buffer was added to the cell-containing coverslip and spread uniformly by placing another coverslip on top of the sample. Samples were then incubated in a humid chamber inside a 37 °C-hybridization oven for 18-36 hours. Encoding hybridization buffer is composed of encoding wash buffer supplemented with 1 mg/mL yeast tRNA (Life Technologies, 15401-011) and 10% w/v dextran sulfate (Sigma, D8906-50G).

Cells were then washed with primary encoding wash buffer, incubated at 47 °C for 10 minutes, and this wash was repeated for a total of three times. A 1:1000 dilution of 0.2-micrometer-diameter carboxylate-modified orange fluorescent beads (Life Technologies, F-8809) in 2xSSC was sonicated for 3 minutes and then incubated with the sample for 5

minutes. The beads were used as fiducial markers to align images obtained from multiple successive rounds of hybridization, as described below. The sample was washed once with 2xSSC, and then post-fixed with 4% v/v paraformaldehyde in 2xSSC at room temperature for 30 minutes. The sample was then washed three times with 2xSSC and either imaged immediately or stored for no longer than 12 hours at 4 °C prior to imaging. All solutions were prepared as RNase-free.

MERFISH imaging with multiple successive rounds of hybridization. The sample coverslip was assembled into a Bioptech's FCS2 flow chamber, and the flow through this chamber was controlled via a home-built fluidics system composed of three computer-controlled 8-way valves (Hamilton, MVP and HVXM 8-5) and a computer-controlled peristaltic pump (Rainin, Dynamax RP-1). The sample was imaged on a home-built microscope constructed around an Olympus IX-71 body and a 1.45 NA, 100x oil immersion objective and configured for oblique incidence excitation. The objective was heated to 37 °C with a Bioptechs objective heater. Constant focus was maintained throughout the imaging process with a home-built, auto-focusing system. Illumination was provided at 641 nm, 561 nm, and 405 nm using solid state lasers (MPB communications, VFL-P500-642; Coherent, 561-200CWCDRH; and Coherent, 1069413/AT) for excitation of the Cy5-labeled readout probes, the fiducial beads, and nuclear counterstains, respectively. These lines were combined with a custom dichroic (Chroma, zy405/488/561/647/752RP-UF1) and the emission was filtered with a custom dichroic (Chroma, ZET405/488/561/647-656/752m). Fluorescence was separated with a QuadView (Photometrics) using the dichroics T560lpxr, T650lpxr, 750dcxxr (Chroma) and the emission filters ET525/50m, WT59550m-2f, ET700/75m, HQ770lp (Chroma) and imaged with an EMCCD camera (Andor, iXon-897). The camera was configured so that a pixel corresponds to 167 nm in the sample plane. The entire system was fully automated, so that imaging and fluid handling were performed for the entire experiment without user intervention.

Sequential hybridization, imaging, and bleaching proceeded as follows. 1 mL of 10 nM of the appropriate fluorescently labeled readout probe in readout hybridization buffer (2xSSC; 10% v/v formamide; 10% w/v dextran sulfate, and 2 mM vanadyl ribonucleoside complex) was flown across the sample, flow was stopped, and the sample was incubated for 15 minutes. Then 2 mL of readout wash buffer (2xSSC, 20% v/v formamide; and 2 mM vanadyl ribonucleoside complex) was flown across the sample, flow was stopped, and the sample was incubated for 3 minutes. 2 mL of imaging buffer comprising 2xSSC, 50 mM TrisHCl pH 8, 10% w/v glucose, 2 mM Trolox (Sigma-Aldrich, 238813), 0.5 mg/mL glucose

oxidase (Sigma-Aldrich, G2133), and 40 microgram/mL catalase (Sigma-Aldrich, C30) was flown across the sample. Flow was then stopped, and then approximately 75 to 100 regions were exposed to ~25 mW 642-nm and 1 mW of 561-nm light and imaged. Each region was 40 micrometers by 40 micrometers. The laser powers were measured at the microscope backport. Because the imaging buffer is sensitive to oxygen, the ~50 mL of imaging buffer used for a single experiment was made fresh at the beginning of the experiment and then stored under a layer of mineral oil throughout the measurement. Buffer stored in this fashion was stable for more than 24 hours.

After imaging, the fluorescence of the readout probes was extinguished via photobleaching. The sample was washed with 2 mL of photobleaching buffer (2xSSC and 2 mM vanadyl ribonucleoside complex), and each imaged region of the sample was exposed to 200 mW of 641-nm light for 3 s. To confirm the efficacy of this photobleaching treatment, imaging buffer was reintroduced, and the sample was imaged as described above.

The above hybridization, imaging, and photobleaching process was repeated either 16 times for the 140-gene measurements using the MHD4 code or 14 times for the 1001-gene measurements using the MHD2 code. An entire experiment was typically completed in ~20 hours.

Following completion of imaging, 2 mL of a 1:1000 dilution of Hoescht (ENZ-52401) in 2xSSC was flown through the chamber to label the nuclei of the cells. The sample was then washed immediately with 2 mL of 2xSSC followed by 2 mL of imaging buffer. Each region of the sample was then imaged once again with ~1 mW of 405-nm light.

Because cells were imaged using wide-field imaging with oblique-incidence illumination, without optical sectioning and z-scanning, the fraction of individual RNA species that was outside the axial range of the imaging geometry was quantified for 6 different RNA species using conventional smFISH. For this purpose, these cells were optically sectioned by collecting stacks of images at different focal depths through the entire depth of the cells. The images were aligned in consecutive focal planes and then computed for each cell the fraction of RNAs that were detected in the three-dimensional stack but not in the basal focal plane. It was found that only a small fraction, 15% +/- 1% (Mean +/- SEM across six different RNA species) of RNA molecules were outside the imaging range of a fixed focal plane without z-scanning. These measurements also confirmed that the excitation geometry illuminated the full depth of the cells. Any optical sectioning technique could be employed in MERFISH to allow the imaging of RNAs in thicker cells or tissues.

Construction of measured words. Fluorescent spots were identified and localized in each image using a multi-Gaussian-fitting algorithm assuming a Gaussian with a uniform width of 167 nm. This algorithm was used to allow partially overlapping spots to be distinguished and individually fit. RNA spots were distinguished from background signal, i.e. signal arising from probes bound non-specifically, by setting the intensity threshold required to fit a spot with this software. Due to variation in the brightness of spots between rounds of hybridization, this threshold was adjusted appropriately for each hybridization round to minimize the combined average of the 1-->0 and 0-->1 error rates across all hybridization rounds (140-gene measurements) or to maximize the ratio of the number of measured words with four '1' bits to those with three or five '1' bits (1001-gene measurements). The location of the fiducial beads was identified in each frame using a faster single-Gaussian fitting algorithm.

Images of the same sample region in different rounds of hybridization were registered by rotating and translating the image to align the two fiducial beads within the same image that were most similar in location after a coarse initial alignment via image correlation. All images were aligned to a coordinate system established by the images collected in the first round of hybridization. The quality of this alignment was determined from the residual distance between five additional fiducial beads, and alignment error was typically ~20 nm.

Fluorescence spots in different hybridization rounds were connected into a single string, corresponding to a potential RNA molecule, if the distance between spots was smaller than 1 pixel (167 nm). For each string of spots, the on-off sequence of fluorescent signals in all hybridization rounds were used to assign a binary word to the potential RNA molecule, in which '1' was assigned to the hybridization rounds that contained a fluorescent signal above threshold and '0' was assigned to the other hybridization rounds. Measured words were then decoded into RNA species using the 16-bit MHD4 code or the 14-bit MHD2 code discussed above. In the case of the 16-bit MHD4 code, if the measured binary word matched the code word of a specific RNA perfectly or differed from the code word by one single bit, it was assigned to that RNA. In the case of the 14-bit MHD2 code, only if the measured binary word matched the code word of a specific RNA perfectly, was it assigned to that RNA. To determine the copy number per cell, the number of each RNA species was counted in individual cells within each 40 micrometer by 40 micrometer imaging area. It is noted that this number accounts for the majority but not all RNA molecules within a cell because a fraction of the cell could be outside the imaging area or focal depth. Tiling images of

adjacent areas and adjacent focal planes could be employed to improve the counting accuracy.

In the 140-gene experiments, some regions of the cell nucleus occasionally contained too much fluorescence signal to properly identify individual RNA spots. In the 1001-gene experiments, the cell nucleus generally contained too much fluorescent signal to allow identification of individual RNA molecules. These bright regions were excluded from all subsequent analysis. This work focuses on mRNAs, which are enriched in the cytoplasm. To estimate the fraction of mRNAs missed by excluding the nucleus region, conventional smFISH was used to quantify the fraction of molecules found inside the nucleus for six different mRNAs species. It was found that only 5% +/- 2% (Mean +/- SEM across six RNA species) of these RNA molecules are found in the nucleus. Employment of super-resolution imaging and/or optical sectioning could potentially allow individual molecules in these dense nucleus regions to be identified, which would be particularly useful for probing those non-coding RNAs that are enriched in the nucleus.

smFISH measurements of individual genes. Pools of 48 fluorescently-labeled (Quasar 670) oligonucleotide probes per RNA were purchased from Biosearch Technologies. 30-nt probe sequences were taken directly from a random subset of the targeting regions used for the multiplexed measurements. Cells were fixed and permeabilized as described above. 10 microliters of 250 nM oligonucleotide probes in encoding hybridization buffer (described above) was added to the cell-containing coverslip and spread uniformly by placing another coverslip on top of the sample. Samples were then incubated in a humid chamber inside a 37 °C-hybridization oven for 18 hours. Cells were then washed with encoding wash buffer (described above) at 37 °C for 10 minutes, and this wash was repeated for a total of three times. The sample was then washed three times with 2xSSC and imaged in imaging buffer using the same imaging geometry as described above for MERFISH.

Bulk RNA sequencing. Total RNA was extracted from IMR90 cells cultured as above using the Zymo Quick RNA MiniPrep kit (R1054) according to the manufacturer's instructions. polyA RNA was then selected (NEB; E7490), and a sequencing library was constructed using the NEBNext Ultra RNA library preparation kit (NEB; E7530), amplified with custom oligonucleotides, and 150-bp reads were obtained from on a MiSeq. These sequences were aligned to the human genome (Gencode v18) and isoform abundance was computed with cufflinks.

Calculation of the predicted scaling and error properties of different encoding schemes. Analytic expressions were derived for the dependence of the number of possible

code words, the calling rate, and the misidentification rate on *N*. The calling rate is defined as the fraction of RNA molecules that are properly identified. The misidentification rate is defined as the fraction of RNA molecules that are misidentified as a wrong RNA species. For encoding schemes with an error-detection capability, the calling rate and misidentification rate does not add up to 1 because a fraction of the molecules not called properly can be detected as errors and discarded and, hence, not misidentified as a wrong species. These calculations assume that the probability of misreading bits is constant for all hybridization rounds but differs for the 1-->0 and 0-->1 errors. Experimentally measured average 1-->0 and 0-->1 error rates (10% and 4% respectively) were used for the estimates shown in Figs. 5B-5D. For simplicity, the word corresponding to all '0's was not removed from calculations.

For the simple binary encoding scheme in which all possible *N*-bit binary words are assigned to unique RNA species, the number of possible code words is 2^N. The number of words that could be used to encode RNA is actually $2^N - 1$ because the code word '00...0' does not contain detectable fluorescence in any hybridization round, but for simplicity the word corresponding to all '0's was not removed from subsequent calculations. The error introduced by this approximation is negligible. For any given word with *m* '1's and *N*–*m* '0's the probability of measuring that word without error (the fraction of RNAs that is properly called) is:

$$(1-p_1)^m(1-p_0)^{N-m}, \tag{1}$$

where p_1 is 1-->0 error rate and p_0 is 0-->1 error rate per bit. Because different words in this simple binary encoding scheme can have different numbers of '1' bits, the calling rate for different words will differ if $p_1 \neq p_0$. The average calling rate, reported in Fig. 5C, was determined from the weighted average of the value of Eq. (1) for all words. This weighted average is:

$$\frac{1}{2^N}\sum_{m=0}^{N}\binom{N}{m}(1-p_1)^m(1-p_0)^{N-m}, \tag{2}$$

where $\binom{N}{m}$ is the binomial coefficient and corresponds to the number of words with *m* '1' bits in this encoding scheme. Since in this encoding scheme every error produces a binary word that encodes a different RNA, the average misidentification rate for this encoding scheme, reported in Fig. 5D, follows directly from (2):

$$1-\frac{1}{2^N}\sum_{m=0}^{N}\binom{N}{m}(1-p_1)^m(1-p_0)^{N-m}. \quad (3)$$

To calculate the scaling and error properties of the extended Hamming distance 4 (HD4) code, the generator matrix for the desired number of data bits using standard methods was first created. The generator matrix determines the specific words that are present in a given encoding scheme and was used to directly determine the number of encoded words as a function of the number of bits. In this encoding scheme, the calling rate corresponds to the fraction of words measured without error as well as the fraction of words measured with a single-bit error. For code words with m '1' bits, this fraction is determined by the following expression:

$$(1-p_1)^m(1-p_0)^{N-m}+m\,p_1{}^1(1-p_1)^{m-1}(1-p_0)^{N-m}+(N-m)\,p_0{}^1(1-p_1)^m(1-p_0)^{N-m-1} \quad (4)$$

where the first term is the probability of not making any errors, the second term corresponds to the total probability of making one 1-->0 error at any of the m '1' bits without making any other 0-->1 errors, and the final term corresponds to the total probability of making one 0-->1 error at any of the N-m '0' bits without making any 1-->0 errors. Because the number of '1' bits can differ between words in this encoding scheme, the average calling rate reported in Fig. 5C was computed from a weighted average over Eq. (4) for different values of m. The weight for each term was determined from the number of words that contain m '1' bits as determined from the generator matrix described above.

Because RNA-encoding words are separated by a minimum Hamming distance of 4, at least 4 errors are required to switch one word into another. If error correction is applied, then 3 or 5 errors could also convert one RNA into another. Thus, the misidentification rate from all possible combinations of 3-bit, 4-bit and 5-bit errors was estimated for code words with m '1' bits. Technically, >5-bit errors could also convert one RNA into another, but the probability of making such errors is negligible because of the small per-bit error rate. This expression was approximated with:

$$\begin{aligned}
&\sum_{i=0}^{4}\binom{m}{i}\binom{N-m}{4-i}p_1^i p_0^{4-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(4-i)}+\\
&\sum_{i=0}^{3}\binom{m}{i}\binom{N-m}{3-i}p_1^i p_0^{3-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(3-i)}+\\
&\sum_{i=0}^{5}\binom{m}{i}\binom{N-m}{5-i}p_1^i p_0^{5-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(5-i)}+
\end{aligned} \quad (5)$$

The first sum corresponds to all of the ways in which exactly four mistakes can be made. Similarly, the second and third sums correspond to all of the ways in which exactly three or

five mistakes can be made. Eq. (5) provides an upper bound for the misidentification rate because not all three, four, or five bit errors produce a word that matches or would be corrected to another legitimate word. Again because the number of '1' bits can differ between words, the average misidentification rate reported in Fig. 5D is calculated as a weighted average of Eq. (5) over the number of words that have *m* '1' bits.

To generate the MHD4 code where the number of '1' bits for each code word is set to 4, the HD4 codes were first generated as described above, and then all code words that did not contain four '1's were removed. The calling rate of this code, reported in Fig. 5C, was directly calculated from Eq. (4) but with m = 4 because all code words in this code have four '1' bits. The misidentification rate of this code, reported in Fig. 5D, was calculated by modifying Eq. (5) with the following considerations: (i) the number of '1' bits, m, was set to 4 and (ii) errors that produce words that do not contain three, four, or five '1' bits were excluded. Thus, the expression in Eq. (5) was simplified to

$$\begin{aligned}
&\binom{4}{2}\binom{N-4}{2}p_1^2 p_0^2(1-p_1)^2(1-p_0)^{N-6} + \\
&\binom{4}{1}\binom{N-4}{2}p_1^1 p_0^2(1-p_1)^3(1-p_0)^{N-6} + \binom{4}{2}\binom{N-4}{1}p_1^2 p_0^1(1-p_1)^2(1-p_0)^{N-5} \\
&\binom{4}{2}\binom{N-4}{3}p_1^2 p_0^3(1-p_1)^2(1-p_0)^{N-7} + \binom{4}{3}\binom{N-4}{2}p_1^3 p_0^2(1-p_1)^1(1-p_0)^{N-6}
\end{aligned} \tag{6}$$

Again, this expression is an upper bound on the actual misidentification rate because not all words with four '1's are valid code words.

Estimates of the 1-->0 and 0-->1 error rates for each hybridization round. To compute the probability of misreading a bit at a given hybridization round, the error correcting properties of the MHD4 code were used. Briefly, the probabilities of 1-->0 or 0-->1 errors were derived in the following way. Let the probability of making an error at the *i*th bit, i.e. *i*th hybridization round, be p_i and the actual number of RNA molecules of the given species be *A*, then the number of exact matches for this RNA will be $W_E = A\prod_{i=1}^{16}(1-p_i)$ and the number of one-bit error corrected matches for this RNA corresponding to errors at the *i*th bit will be $W_i = A\frac{p_i}{(1-p_i)}\prod_{j=1}^{16}(1-p_j)$. The p_i can be directly derived from the ratio: $W_i/W_E = \frac{p_i}{(1-p_i)}$. This ratio assumes that the one-bit error-corrected counts were only generated from single-bit errors from the correct word and that multi-error contamination from other RNA words is negligible. Given that the error rate per hybridization round is small and that it takes at least

three errors to convert one RNA-encoding word into a word that would be misidentified as another RNA, the above approximation should be a good one.

To compute the average 1-->0 or 0-->1 error probabilities for each of the 16 hybridization rounds, the above approach was used to calculate the per-bit error rates for each bit of every gene, and these errors were sorted based on whether they correspond to a 1-->0 or a 0-->1 error, and the average of these errors for each bit weighted by the number of counts observed for the corresponding gene was taken.

Estimates of the calling rate for individual RNA species from actual imaging data. With the estimates of the 1-->0 or 0-->1 error probabilities for each round of hybridization as determined above, it is possible to estimate the calling rate for each RNA based on the specific word used to encoded it. Specifically, the fraction of an RNA species that is called correctly is determined by

$$\prod_{i=1}^{N}(1-p_i)+\sum_{j=1}^{N}\frac{p_j}{(1-p_j)}\prod_{i=1}^{N}(1-p_i), \tag{7}$$

where the first term represent the probability of observing an exact match of the code word and the second term represent the probability of observing an error-corrected match (i.e. with one-bit error). The values of the per-bit error rate p_i for each RNA species are determined by the specific code word for that RNA and the measured 1-->0 or 0-->1 error rates for each round of hybridization. If the code word of the RNA contains a '1' in the *i*th bit, then p_i is determined from the 1-->0 error rate for the *i*th hybridization round; if the word contains a '0' in the *i*th bit, p_i is determined from the 0-->1 error rate for the *i*th hybridization round.

Hierarchical clustering analysis of the co-variation in RNA abundance. Hierarchical clustering of the co-variation in gene expression for both the 140-gene and 1001-gene experiments was conducted as follows. First, the distance between every pair of genes was determined as 1 minus the Pearson correlation coefficient of the cell-to-cell variation of the measured copy numbers of these two RNA species, both normalized by the total RNA counted in the cell. Thus, highly correlated genes are 'closer' to one another and highly anti-correlated genes are 'further' apart. An agglomerative hierarchical cluster tree was then constructed from these distances using the Unweighted Pair Group Method with Arithmetic mean (UPGMA). Specifically, starting with individual genes, hierarchical clusters were constructed by identifying the two clusters (or individual genes) that are closest to one another according to the arithmetic mean of the distances between all inter-cluster gene pairs. The pairs of clusters (or individual genes) with the smallest distance are then grouped

together and the process is repeated. The matrix of pairwise correlations was then sorted based on the order of the genes within these trees.

Groups of genes with substantial co-variations were identified by selecting a threshold on the hierarchical cluster tree (indicated by the dashed lines in Figs. 7D and 10A) that produced approximately 10 groups of genes each of which contains at least 4 members for the 140-gene experiments or approximately 100 groups each of which contains at least 3 members for the 1001-gene experiments. It is noted that one can change the threshold in order to identify either more tightly coupled smaller groups or larger groups with relatively loose coupling.

A probability value for the confidence that a gene belongs to a specific group was determined by computing the difference between the average correlation coefficient between that gene and all other members of that group and the average correlation coefficient between that gene and all other measured genes outside that group. The significance (p-value) of this difference was determined with the student's t-test.

Because hierarchical clustering is inherently a one-dimensional analysis, i.e. any given genes can only be a member of a single group, this analysis does not allow all correlated gene groups to be identified. Higher dimension analysis, such as principal component analysis or k-means clustering, could be used to identify more co-varying gene clusters.

Analysis of RNA spatial distributions. To identify genes that have similar spatial distributions, each of the measured cells was subdivided into 2x2 regions and calculated the fraction of each RNA species present in each of these bins. To control for the fact that some regions of the cell naturally contain more RNA than others, the enrichment was calculated for each gene , i.e., the ratio of the observed fraction in a given region for a given RNA species to the average fraction observed for all genes in that same region. For each pair of RNA species, the Pearson correlation coefficient of the region-to-region variation in enrichment of these two RNA species for each cell was determined and the correlation coefficients were averaged over ~400 cells imaged in 7 independent data sets. RNA species were then clustered based on these average correlation coefficients using the same hierarchical clustering algorithm described above. Because of the large number of cells used for the analysis, it was found that the coarse spatial binning (2x2 regions per cell) was sufficient to capture the spatial correlation between genes and finer binning did not produce more significantly correlated groups.

To measure the distances of genes from the nuclei and from the cell edge, brightness thresholds on the cell images were first used to segment the nuclei and the cell edges identified. The distance from every RNA molecule to the nearest part of the nucleus and nearest part of the cell edge was then determined. For each data set, the average distance for each RNA species averaged over all the cells measured was then computed. These distances were averaged for the group I genes, group II genes or all genes. Only those RNA species with at least 10 counts per cell were used in this analysis to minimize statistical error on the distance values.

Gene ontology (GO) analysis. Groups of genes were selected from the hierarchical trees as discussed above. A collection of GO terms was determined for all measured RNA species as well as the RNA species associated with each group from the most recent human GO annotations using both the annotated GO terms and terms immediately upstream or downstream of the found annotations. The enrichment of these annotations was calculated from the ratio of the fraction of genes within each group that have this term to the fraction of all measured genes that have this term and the p-value for this enrichment was calculated via the hypergeometric function. Only statistically significantly enriched GO terms with a p-value less than 0.05 were considered.

While several embodiments of the present invention have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the functions and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the present invention. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the teachings of the present invention is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, the invention may be practiced otherwise than as specifically described and claimed. The present invention is directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles,

materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the scope of the present invention.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

The indefinite articles "a" and "an," as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean "at least one."

The phrase "and/or," as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with "and/or" should be construed in the same fashion, i.e., "one or more" of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the "and/or" clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to "A and/or B", when used in conjunction with open-ended language such as "comprising" can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, "or" should be understood to have the same meaning as "and/or" as defined above. For example, when separating items in a list, "or" or "and/or" shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as "only one of" or "exactly one of," or, when used in the claims, "consisting of," will refer to the inclusion of exactly one element of a number or list of elements. In general, the term "or" as used herein shall only be interpreted as indicating exclusive alternatives (i.e. "one or the other but not both") when preceded by terms of exclusivity, such as "either," "one of," "only one of," or "exactly one of." "Consisting essentially of," when used in the claims, shall have its ordinary meaning as used in the field of patent law.

As used herein in the specification and in the claims, the phrase "at least one," in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also

allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase "at least one" refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, "at least one of A and B" (or, equivalently, "at least one of A or B," or, equivalently "at least one of A and/or B") can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

When the word "about" is used herein in reference to a number, it should be understood that still another embodiment of the invention includes that number not modified by the presence of the word "about."

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

In the claims, as well as in the specification above, all transitional phrases such as "comprising," "including," "carrying," "having," "containing," "involving," "holding," "composed of," and the like are to be understood to be open-ended, i.e., to mean including but not limited to. Only the transitional phrases "consisting of" and "consisting essentially of" shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

Codebook 1 for the 140-gene experiments

Genes	Codewords
THBS1	0 0 0 0 1 1 1 0 0 0 0 0 0 0 1 0
FLNA	0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 1
TNC	1 0 0 0 1 0 1 0 0 0 0 0 1 0 0 0
FBN2	0 0 1 0 1 0 1 1 0 0 0 0 0 0 0 0
KIAA1199	0 0 0 0 1 1 0 1 0 0 0 1 0 0 0 0
COL5A1	0 0 0 0 0 1 0 0 1 1 0 0 0 0 1 0
MALAT1	0 0 0 0 0 1 1 0 1 0 0 0 1 0 0 0
TLN1	0 0 0 0 1 0 0 0 0 0 1 0 0 1 1 0
FBN1	0 0 1 0 0 1 0 1 0 0 0 0 0 0 1 0
SRRM2	0 0 1 0 0 0 1 0 0 0 0 0 0 0 1 1
TEAD1	0 0 1 0 0 1 1 0 0 0 0 1 0 0 0 0
IQGAP1	0 0 0 0 0 0 1 1 0 0 0 1 0 0 1 0
MYH10	0 0 1 0 1 0 0 0 0 0 0 1 0 0 1 0
PAPPA	0 0 0 1 0 1 0 0 0 0 1 0 0 0 1 0
NR4A1	0 0 0 1 1 0 1 0 0 0 1 0 0 0 0 0
LRP1	0 0 0 0 0 1 1 0 0 0 1 0 0 1 0 0
SLC38A1	1 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0
SMARCA5	0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 1
CKAP5	0 1 0 0 0 0 1 0 0 0 1 0 0 0 1 0
PIEZO1	1 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0
VCAN	0 0 0 0 1 0 1 0 1 1 0 0 0 0 0 0
PRRC2B	0 1 0 1 0 1 1 0 0 0 0 0 0 0 0 0
SON	1 0 0 0 0 1 0 0 0 0 0 0 1 0 1 0
SLC7A11	1 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0
GPR107	0 1 0 0 0 1 0 0 0 0 0 0 0 1 1 0
KIAA1462	0 1 0 1 1 0 0 0 0 0 0 0 0 0 1 0
notarget001	0 0 0 0 1 0 0 0 1 0 0 0 1 0 1 0
blank001	0 1 0 0 1 1 0 0 0 0 1 0 0 0 0 0
ROCK2	0 1 0 0 1 0 1 0 0 0 0 0 0 1 0 0
SPTBN1	0 0 0 1 1 1 0 0 0 0 0 0 0 1 0 0
FLNC	0 0 0 1 0 0 1 0 0 0 0 0 0 1 1 0
PRPF8	0 0 0 0 0 1 0 0 0 0 0 1 0 0 1 1
IGF2R	0 0 0 0 1 1 0 0 0 1 0 0 1 0 0 0
FOSB	1 0 0 0 0 1 1 0 0 1 0 0 0 0 0 0
UBR5	0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 1
SVEP1	0 0 0 0 0 0 1 0 0 1 0 0 1 0 1 0
CBX5	0 0 1 0 1 1 0 0 0 0 0 0 0 0 0 1
PDS5A	1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 1
NUMA1	0 0 0 0 0 0 1 0 0 1 1 0 0 0 0 1
PLXNA1	0 0 0 1 0 0 0 0 0 0 0 1 1 0 1 0

Fig. 20A

Codebook 1 for the 140-gene experiments

Genes	Codewords
ALPK2	0 1 0 0 0 1 0 0 0 1 0 1 0 0 0 0
AFAP1	0 0 0 0 0 0 1 0 1 0 1 1 0 0 0 0
SPTAN1	0 0 1 1 1 0 0 0 0 0 0 0 1 0 0 0
DOCK7	0 1 1 0 0 0 0 0 0 0 0 0 1 0 1 0
NOTCH2	0 0 0 0 0 0 0 0 1 0 0 0 0 1 1 1
DYNC1H1	1 0 0 0 1 0 0 0 0 0 0 1 0 1 0 0
ITGA2	1 0 1 0 0 0 0 0 0 0 0 0 1 1 0
PRKDC	0 0 0 1 1 0 0 0 0 1 0 0 0 0 0 1
GTF3C1	0 0 1 0 0 1 0 0 0 0 0 0 1 1 0 0
THSD4	0 0 0 0 0 1 0 0 0 1 0 0 0 1 0 1
PTPN14	0 1 0 0 1 0 0 1 0 1 0 0 0 0 0 0
LUZP1	0 0 0 0 1 0 0 0 1 0 1 0 0 0 0 1
AMOTL1	1 0 0 0 0 0 1 1 0 0 1 0 0 0 0 0
MAN1A2	1 0 1 0 1 0 0 0 0 0 1 0 0 0 0 0
notarget002	1 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0
blank002	1 1 0 0 0 0 0 1 0 0 0 0 0 0 1 0
USP9X	0 0 0 0 1 0 0 0 0 1 1 1 0 0 0 0
CREBBP	0 1 0 0 0 0 1 0 0 0 0 0 1 0 0 1
KPNA4	0 1 1 0 0 0 1 0 0 1 0 0 0 0 0 0
EGFR	0 0 0 1 0 0 1 0 0 1 0 1 0 0 0 0
ZC3H13	0 0 0 1 0 0 1 1 0 0 0 0 1 0 0 0
TNRC6A	0 0 0 1 0 1 0 1 1 0 0 0 0 0 0 0
TPR	0 0 0 0 0 1 0 0 1 0 0 1 0 1 0 0
COL7A1	0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 1
ASCC3	0 1 0 0 0 0 1 1 1 0 0 0 0 0 0 0
USP34	0 0 0 0 0 0 0 0 0 0 1 0 1 0 1 1
CENPF	0 0 1 0 0 0 1 0 1 0 0 0 0 1 0 0
KIDINS220	0 0 0 0 0 1 0 0 0 0 1 1 1 0 0 0
SCUBE3	1 0 0 0 0 1 0 1 0 0 0 0 0 1 0 0
NRIP1	0 0 0 0 0 0 0 0 0 1 0 1 0 1 1 0
NUP98	1 0 0 1 0 0 0 0 0 0 0 0 0 0 1 1
ZFC3H1	0 0 0 0 0 0 0 1 1 0 1 0 0 0 1 0
FAM208B	0 1 0 0 0 0 0 0 0 1 0 0 0 0 1 1
RAB3B	0 0 0 1 0 0 0 1 0 1 0 0 0 0 1 0
AFF4	1 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0
PRKCA	0 0 1 0 0 0 1 0 0 0 1 0 1 0 0 0
FLNB	1 0 0 0 0 0 1 0 0 0 0 0 0 1 0 1
CHD8	1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1
ANKRD52	0 1 0 0 1 0 0 0 0 0 0 1 1 0 0 0
MKI67	0 0 0 0 0 0 1 0 0 0 0 1 1 1 0 0

Fig. 20B

Codebook 1 for the 140-gene experiments

Genes	Codewords
DIP2C	0 1 1 0 1 0 0 0 1 0 0 0 0 0 0 0
SSH1	0 0 0 0 1 0 0 0 0 0 0 0 1 1 0 1
notarget003	0 0 1 0 0 1 0 0 1 0 1 0 0 0 0 0
blank003	1 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0
DNAJC13	0 0 0 0 1 0 0 1 1 0 0 0 0 1 0 0
DSEL	0 1 0 0 0 0 0 0 1 0 0 1 0 0 1 0
PIK3CA	0 0 0 0 0 1 0 1 0 1 1 0 0 0 0 0
MED14	1 0 0 1 0 1 0 0 0 0 0 1 0 0 0 0
HERC2	0 1 0 0 0 1 0 0 1 0 0 0 0 0 0 1
H6PD	0 0 1 1 0 0 0 0 1 0 0 0 0 0 1 0
C14orf132	0 0 0 0 0 0 0 1 0 0 0 0 1 1 1 0
SIPA1L3	0 0 0 0 0 0 0 1 1 0 1 0 0 0 1 0 0
C17orf51	0 0 1 1 0 1 0 0 0 1 0 0 0 0 0 0
NF1	0 0 0 1 0 0 1 0 1 0 0 0 0 0 0 1
VPS13A	1 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0
ABCA2	1 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0
PCNX	0 1 0 0 0 1 0 1 0 0 0 0 1 0 0 0
NKTR	0 0 1 0 1 0 0 0 0 1 0 0 0 1 0 0
FYCO1	0 0 0 1 1 0 0 0 1 0 0 1 0 0 0 0
ATP11A	0 0 1 0 0 0 0 0 0 1 1 0 0 0 1 0
REV3L	0 0 0 0 1 0 0 1 0 0 1 0 1 0 0 0
CASC5	0 1 0 0 0 0 0 1 0 0 1 1 0 0 0 0
AKAP11	0 1 0 1 0 0 0 0 0 0 1 0 0 1 0 0
CEP250	0 0 1 0 0 0 0 0 0 1 0 0 1 0 0 1
USP24	0 0 0 0 0 0 0 0 1 0 1 0 1 1 0 0
AGPS	1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1
CBL	1 0 1 0 0 0 0 0 0 1 0 1 0 0 0 0
RNF169	0 0 0 1 0 0 0 0 0 0 1 1 0 0 0 1
TMOD2	0 0 0 1 0 0 0 0 0 1 0 0 1 1 0 0
FASN	0 0 0 0 0 0 0 0 1 1 0 1 0 0 0 1
notarget004	1 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0
blank004	0 0 1 0 0 0 0 0 1 0 0 1 1 0 0 0
UBR2	0 1 1 0 0 0 0 0 0 0 1 0 0 0 0 1
PHIP	1 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1
GTF3C4	0 0 0 0 0 0 0 1 0 1 0 1 1 0 0 0
QSER1	1 0 0 1 0 0 0 0 0 0 1 0 1 0 0 0
VPS13D	0 0 1 0 0 0 0 1 1 1 0 0 0 0 0 0
FRY	0 0 0 1 0 0 0 0 0 1 1 1 0 0 0 0 0
DIP2B	0 1 0 1 0 0 0 0 0 1 0 0 0 1 0 0 0
HIVEP2	1 1 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0

Fig. 20C

Codebook 1 for the 140-gene experiments

Genes	Codewords
DOPEY1	1 0 0 0 0 0 0 0 1 1 0 0 1 0 0 0
ALMS1	0 1 1 0 0 0 0 1 0 0 0 0 0 1 0 0
HEATR5B	0 0 1 1 0 0 0 1 0 0 1 0 0 0 0 0
HIVEP1	0 0 1 0 0 0 0 1 0 0 0 1 0 0 0 1
ARL10	0 1 0 1 0 0 0 1 0 0 0 0 0 0 0 1
HELZ2	0 0 0 0 0 0 0 1 1 0 0 0 1 0 0 1
POLQ	1 0 1 0 0 0 0 0 1 0 0 0 0 0 0 1
CACNA1A	1 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0
KIAA2018	0 1 0 0 0 0 0 0 0 0 0 0 1 0 1 0 1
LYST	1 0 0 0 0 0 0 1 1 0 0 1 0 0 0 0
ZKSCAN2	0 0 1 1 0 0 0 0 0 0 0 0 0 1 0 1
DIEXF	0 0 0 1 0 0 0 1 0 0 0 1 0 1 0 0
FZD4	1 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0
SLC5A3	0 1 1 1 0 0 0 0 0 0 0 1 0 0 0 0
KIAA1147	1 1 0 0 0 0 0 0 1 0 1 0 0 0 0 0
LMTK2	0 1 0 0 0 0 0 0 0 0 1 1 0 1 0 0 0
STARD9	0 0 1 0 0 0 0 0 0 0 0 1 1 0 1 0 0
BRCA2	0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 1
notarget005	1 0 0 0 0 0 0 0 0 1 1 0 0 1 0 0
blank005	1 1 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0

Fig. 20D

Codebook 2 (shuffled) for the 140-gene experiments

Genes	Codewords
THBS1	0 0 0 0 1 0 0 0 0 0 0 0 0 1 1 0 1
FLNA	0 0 0 0 1 1 0 0 0 1 0 0 1 0 0 0
TNC	1 0 0 0 0 0 1 0 0 0 0 0 0 1 0 1
FBN2	0 1 0 0 0 0 0 0 0 0 0 1 0 1 0 1
KIAA1199	0 0 0 0 0 0 0 0 0 0 1 0 1 0 1 1
COL5A1	0 0 0 0 0 0 0 1 0 0 0 0 1 1 1 0
MALAT1	0 0 0 0 0 0 0 0 1 0 0 0 0 1 1 1
TLN1	0 0 0 1 0 0 0 0 0 1 0 0 1 1 0 0
FBN1	1 0 0 0 1 0 0 0 0 0 0 1 0 1 0 0
SRRM2	0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 1
TEAD1	1 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0
IQGAP1	0 0 1 0 1 1 0 0 0 0 0 0 0 0 0 1
MYH10	0 0 0 1 1 0 0 0 0 1 0 0 0 0 0 1
PAPPA	0 0 0 0 1 0 0 1 0 0 1 0 1 0 0 0
NR4A1	0 0 0 0 1 0 0 0 1 0 1 0 0 0 0 1
LRP1	0 0 0 0 0 0 0 0 1 0 1 0 1 1 0 0
SLC38A1	0 0 0 0 0 1 0 0 0 1 0 0 0 1 0 1
SMARCA5	0 0 1 0 0 0 0 0 0 1 0 0 1 0 0 1
CKAP5	0 0 1 0 0 1 0 0 0 0 0 0 1 1 0 0
PIEZO1	0 0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 1
VCAN	0 0 0 0 0 0 0 1 0 0 0 0 1 1 1 0 0
PRRC2B	1 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1
SON	0 0 1 0 1 0 0 0 0 1 0 0 0 1 0 0
SLC7A11	0 1 0 0 1 0 1 0 0 0 0 0 0 1 0 0
GPR107	0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 1
KIAA1462	0 0 1 1 1 0 0 0 0 0 0 0 1 0 0 0
notarget001	1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 1
blank001	1 0 0 0 1 0 1 0 0 0 0 0 1 0 0 0
ROCK2	0 1 0 0 0 0 1 0 0 0 0 0 1 0 0 1
SPTBN1	0 0 0 0 1 0 0 0 0 0 1 0 0 1 1 0
FLNC	0 0 0 0 1 0 0 0 1 0 0 0 1 0 1 0
PRPF8	0 0 0 0 0 0 0 1 1 0 0 0 1 0 0 1
IGF2R	0 0 0 1 1 1 0 0 0 0 0 0 0 1 0 0
FOSB	0 0 1 1 0 0 0 0 0 0 0 0 0 1 0 1
UBR5	0 1 0 0 1 0 0 0 0 0 0 1 1 0 0 0
SVEP1	0 0 0 0 1 0 0 1 1 0 0 0 0 1 0 0
CBX5	0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 1
PDS5A	1 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0
NUMA1	0 1 0 0 0 1 0 1 0 0 0 0 1 0 0 0
PLXNA1	0 1 0 1 0 0 0 0 0 0 1 0 0 1 0 0

Fig. 20E

Codebook 2 (shuffled) for the 140-gene experiments

Genes	Codewords
ALPK2	0 1 1 0 0 0 0 0 0 0 0 0 0 1 0 1 0
AFAP1	1 0 0 0 0 0 0 0 0 1 1 0 0 1 0 0
SPTAN1	0 0 0 0 0 0 0 0 0 1 0 1 0 1 1 0
DOCK7	0 0 1 0 0 0 0 0 0 0 1 1 0 1 0 0
NOTCH2	0 0 0 0 0 1 1 0 1 0 0 0 1 0 0 0
DYNC1H1	0 1 1 0 1 0 0 0 1 0 0 0 0 0 0 0
ITGA2	0 0 0 1 1 0 0 0 1 0 0 1 0 0 0 0
PRKDC	0 0 0 0 0 0 0 0 1 1 0 1 0 0 0 1
GTF3C1	0 0 1 0 1 0 0 0 0 0 0 1 0 0 1 0
THSD4	1 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1
PTPN14	0 0 0 0 0 0 1 1 0 1 0 0 0 1 0 0
LUZP1	1 0 1 0 0 0 0 0 0 0 0 0 0 1 1 0
AMOTL1	0 0 0 0 1 0 1 0 1 1 0 0 0 0 0 0
MAN1A2	0 1 0 1 0 0 0 1 0 0 0 0 0 0 0 1
notarget002	1 0 0 0 0 1 0 0 0 0 0 0 1 0 1 0
blank002	1 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0
USP9X	0 0 0 0 0 0 0 1 0 1 0 1 1 0 0 0
CREBBP	0 1 0 0 0 1 0 0 1 0 0 0 0 0 0 1
KPNA4	1 0 0 1 0 0 0 0 0 0 0 0 0 0 1 1
EGFR	1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1
ZC3H13	1 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0
TNRC6A	0 0 0 0 1 0 0 0 0 1 1 1 0 0 0 0
TPR	1 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0
COL7A1	0 0 1 0 0 0 0 1 0 0 0 1 0 0 0 1
ASCC3	0 1 0 0 1 0 0 1 0 1 0 0 0 0 0 0
USP34	0 0 1 0 0 0 1 0 0 0 1 0 1 0 0 0
CENPF	0 0 1 0 0 0 0 0 1 0 0 1 1 0 0 0
KIDINS220	0 0 0 1 0 0 0 0 0 0 1 1 0 0 0 1
SCUBE3	0 0 0 0 0 0 1 0 0 1 1 0 0 0 0 1
NRIP1	0 0 0 1 1 0 1 0 0 0 1 0 0 0 0 0
NUP98	0 0 0 0 0 1 0 0 0 0 1 1 1 0 0 0
ZFC3H1	0 0 0 1 0 0 1 0 1 0 0 0 0 0 0 1
FAM208B	0 1 0 1 1 0 0 0 0 0 0 0 0 0 1 0
RAB3B	0 0 0 1 0 0 0 0 0 0 0 0 1 1 0 1 0
AFF4	0 1 0 0 0 0 0 0 0 1 0 0 0 0 1 1
PRKCA	0 1 0 0 0 1 0 0 0 0 0 0 0 1 1 0
FLNB	0 1 0 0 0 0 0 0 1 1 0 0 0 1 0 0
CHD8	0 0 0 0 0 1 0 0 0 0 0 1 0 0 1 1
ANKRD52	0 0 1 0 0 0 1 0 0 0 0 0 0 0 1 1
MKI67	0 1 0 0 1 1 0 0 0 0 1 0 0 0 0 0

Fig. 20F

Codebook 2 (shuffled) for the 140-gene experiments

Genes	Codewords
DIP2C	1 0 1 0 1 0 0 0 0 0 1 0 0 0 0 0
SSH1	0 1 0 1 0 0 0 0 1 0 0 0 1 0 0 0
notarget003	0 1 1 0 0 0 0 0 0 0 1 0 0 0 0 1
blank003	0 0 0 0 1 1 1 0 0 0 0 0 0 0 1 0
DNAJC13	0 0 0 1 0 0 1 1 0 0 0 0 1 0 0 0
DSEL	0 0 0 1 0 0 0 1 0 0 0 1 0 1 0 0
PIK3CA	0 0 0 0 0 1 1 0 0 0 1 0 0 1 0 0
MED14	0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 1
HERC2	1 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0
H6PD	0 0 1 0 0 0 1 0 1 0 0 0 0 1 0 0
C14orf132	0 0 0 0 0 1 0 0 1 0 0 1 0 1 0 0
SIPA1L3	1 0 1 0 0 0 0 0 1 0 0 0 0 0 0 1
C17orf51	0 1 0 0 0 0 0 0 0 1 1 0 1 0 0 0
NF1	1 0 0 0 0 0 0 0 1 1 0 0 1 0 0 0
VPS13A	0 0 1 0 1 0 1 1 0 0 0 0 0 0 0 0
ABCA2	0 0 0 0 0 0 1 0 0 1 0 0 1 0 1 0
PCNX	0 0 0 1 0 0 1 0 0 0 0 0 0 1 1 0
NKTR	0 1 1 0 0 0 0 1 0 0 0 0 0 1 0 0
FYCO1	0 0 0 0 1 1 0 1 0 0 0 1 0 0 0 0
ATP11A	1 0 0 1 0 0 0 0 0 0 1 0 1 0 0 0
REV3L	1 0 0 0 0 1 0 1 0 0 0 0 0 1 0 0
CASC5	0 0 0 0 0 1 0 1 0 1 1 0 0 0 0 0
AKAP11	0 0 1 0 0 1 0 0 1 0 1 0 0 0 0 0
CEP250	1 0 1 0 0 0 0 0 0 1 0 1 0 0 0 0
USP24	1 1 0 1 0 0 0 0 0 1 0 0 0 0 0 0
AGPS	1 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0
CBL	0 1 0 0 0 0 1 1 1 0 0 0 0 0 0 0
RNF169	0 0 1 1 0 0 0 1 0 0 1 0 0 0 0 0
TMOD2	0 0 1 0 0 1 0 1 0 0 0 0 0 0 1 0
FASN	0 0 0 0 0 0 1 1 0 0 0 1 0 0 1 0
notarget004	0 1 1 0 0 0 1 0 0 1 0 0 0 0 0 0
blank004	1 1 0 0 0 0 0 1 0 0 0 0 0 0 1 0
UBR2	0 1 0 0 0 0 0 0 1 0 0 1 0 0 1 0
PHIP	0 0 1 0 0 1 1 0 0 0 0 1 0 0 0 0
GTF3C4	0 1 0 0 0 0 1 0 0 0 1 0 0 0 1 0
QSER1	1 0 0 0 0 0 1 1 0 0 1 0 0 0 0 0
VPS13D	1 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0
FRY	1 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0
DIP2B	0 1 0 0 0 1 0 0 0 1 0 1 0 0 0 0
HIVEP2	1 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0

Fig. 20G

Codebook 2 (shuffled) for the 140-gene experiments

Genes	Codewords
DOPEY1	0 0 1 1 0 0 0 0 1 0 0 0 0 0 1 0
ALMS1	0 0 1 1 0 1 0 0 0 1 0 0 0 0 0 0
HEATR5B	0 0 1 0 0 0 0 1 1 1 0 0 0 0 0 0
HIVEP1	0 0 0 1 0 1 0 0 0 0 1 0 0 0 1 0
ARL10	0 0 0 0 0 1 0 0 1 1 0 0 0 0 1 0
HELZ2	0 0 0 1 0 0 0 1 0 1 0 0 0 0 1 0
POLQ	0 0 0 0 0 0 1 0 1 0 1 1 0 0 0 0
CACNA1A	1 0 0 0 0 1 1 0 0 1 0 0 0 0 0 0
KIAA2018	0 0 0 0 0 0 0 1 1 0 1 0 0 0 1 0
LYST	0 0 0 1 0 1 0 1 1 0 0 0 0 0 0 0
ZKSCAN2	0 1 0 1 0 1 1 0 0 0 0 0 0 0 0 0
DIEXF	0 1 0 0 0 0 0 1 0 0 1 1 0 0 0 0
FZD4	0 0 0 1 0 0 1 0 0 1 0 1 0 0 0 0
SLC5A3	1 0 0 0 0 0 0 1 1 0 0 1 0 0 0 0
KIAA1147	0 0 1 0 0 0 0 0 0 1 1 0 0 0 1 0
LMTK2	0 1 1 1 0 0 0 0 0 0 0 1 0 0 0 0
STARD9	0 0 0 1 0 0 0 0 1 1 1 0 0 0 0 0
BRCA2	1 0 0 1 0 1 0 0 0 0 0 1 0 0 0 0
notarget005	1 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0
blank005	1 1 0 0 0 0 0 0 1 0 1 0 0 0 0 0

Fig. 20H

Electronic Patent Application Fee Transmittal

Application Number:	15329683
Filing Date:	27-Jan-2017
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Filer:	Tani Chen/Eileen MacKenzie
Attorney Docket Number:	H0498.70509US02

Filed as Small Entity

Filing Fees for U.S. National Stage under 35 USC 371

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension - 1 month with $0 paid	2251	1	100	100
Miscellaneous:				
		Total in USD ($)		**100**

Electronic Acknowledgement Receipt

EFS ID:	36721221
Application Number:	15329683
International Application Number:	
Confirmation Number:	7428
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Customer Number:	86110
Filer:	Tani Chen/Eileen MacKenzie
Filer Authorized By:	Tani Chen
Attorney Docket Number:	H0498.70509US02
Receipt Date:	29-JUL-2019
Filing Date:	27-JAN-2017
Time Stamp:	19:29:31
Application Type:	U.S. National Stage under 35 USC 371

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$100
RAM confirmation Number	E20197SJ29533620
Deposit Account	232825
Authorized User	Eileen MacKenzie
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows: 37 CFR 1.17 (Patent application and reexamination processing fees)	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	H049870509US02-TRN-TC.pdf	26274 0ae78572e4b3e7da2ac9d2c07cf276ce74d0ea6f	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	H049870509US02-FEE-TC.pdf	32664 a6b8cdc7d841fe51b7822533a12abed50ac214a0	no	1
Warnings:					
Information:					
3	Extension of Time	H049870509US02-EXT-TC.pdf	24610 e720d9cf1a756f457eeb0be2ccc4573eec9fff1c	no	1
Warnings:					
Information:					
4		H049870509US02-AMN-TC.pdf	56005 5d5f7f761197e314534418bc4282a7264866b84e	yes	12

Multipart Description/PDF files in .zip description

Document Description	Start	End
Amendment/Req. Reconsideration-After Non-Final Reject	1	1
Specification	2	4
Claims	5	9
Applicant Arguments/Remarks Made in an Amendment	10	12

Warnings:

Information:

5	Specification	H049870509US02-MUSPEC-TC.pdf	347860 2fff551198af2ba8db4e7588fa9fad74d92f02d3	no	90
Warnings:					
Information:					
6	Specification	H049870509US02-SUBSPEC-TC.pdf	347592 80ef9972349607636e250e79a4c8b4d3162c233d	no	90
Warnings:					
Information:					
7	Drawings-other than black and white line drawings	H049870509US02-RDRW-TC.pdf	571170 293428637254534 8d2a03a1c5af227a11b229fb8	no	8
Warnings:					
Information:					
8	Fee Worksheet (SB06)	fee-info.pdf	30931 d6cef2665a77ab09fd107832cc3fe430a98fde15	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	1437106		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/329,683-Conf. #7428
Filing Date	January 27, 2017
First Named Inventor	Xiaowei Zhuang
Art Unit	1634
Examiner Name	F. W. Lu
Attorney Docket Number	H0498.70509US02

ENCLOSURES (*Check all that apply*)

[x] Fee Transmittal Form	[ ] Drawing(s)	[ ] After Allowance Communication to TC
[ ] Fee Attached	[ ] Licensing-related Papers	[ ] Appeal Communication to Board of Appeals and Interferences
[x] Amendment/Reply	[ ] Petition	[ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
[ ] After Final	[ ] Petition to Convert to a Provisional Application	[ ] Proprietary Information
[ ] Affidavits/declaration(s)	[ ] Power of Attorney, Revocation Change of Correspondence Address	[ ] Status Letter
[x] Extension of Time Request	[ ] Terminal Disclaimer	[x] Other Enclosure(s) (please Identify below):
[ ] Express Abandonment Request	[ ] Request for Refund	- Substitute Specification (marked-up and clean versions)
[ ] Information Disclosure Statement	[ ] CD, Number of CD(s) _____	- Replacement Figures (8 sheets)
[ ] Certified Copy of Priority Document(s)	[ ] Landscape Table on CD	
[ ] Reply to Missing Parts/ Incomplete Application	Remarks	
[ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53		

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Tani Chen/		
Printed name	Tani Chen, Sc.D.		
Date	July 29, 2019	Reg. No.	52,728

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: July 29, 2019 Electronic Signature for : /Eileen M. MacKenzie/

7125334.1

FEE TRANSMITTAL

Complete if known	
Application Number	15/329,683-Conf. #7428
Filing Date	January 27, 2017
First Named Inventor	Xiaowei Zhuang
Examiner Name	F. W. Lu
Art Unit	1634
Practitioner Docket No.	H0498.70509US02

[X] Applicant asserts small entity status. See 37 CFR 1.27.

[ ] Applicant certifies micro entity status. See 37 CFR 1.29. Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

TOTAL AMOUNT OF PAYMENT ($)100.00

METHOD OF PAYMENT (check all that apply)

[ ] Check [X] Credit Card [ ] Money Order [ ] None [ ] Other (please identify): _____

[X] Deposit Account Deposit Account Number: 23/2825 Deposit Account Name: Wolf, Greenfield & Sacks, P.C.

For the above-identified deposit account, the Director is hereby authorized to (check all that apply):

[X] Charge fee(s) indicated below

[ ] Charge fee(s) indicated below, **except for the filing fee**

[X] Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17, **except for any excess claims fees or multiple dependent claim fee**

[ ] Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17

[X] Credit any overpayment of fee(s)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES (U = undiscounted fee; S = small entity fee; M = micro entity fee)

	FILING FEES			SEARCH FEES			EXAMINATION FEES			
Application Type	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**Fees Paid ($)**
Utility	300	150*	75	660	330	165	760	380	190	
Design	200	100	50	160	80	40	600	300	150	
Plant	200	100	50	420	210	105	620	310	155	
Reissue	300	150	75	660	330	165	2,200	1,100	550	
Provisional	280	140	70	0	0	0	0	0	0	

* The $150 small entity status filing fee for a utility application is further reduced to $75 for a small entity status applicant who files the application via EFS-Web.

2. EXCESS CLAIM FEES

Fee Description	**Undiscounted Fee ($)**	**Small Entity Fee ($)**	**Micro Entity Fee ($)**
Each claim over 20 (including Reissues)	100	50	25
Each independent claim over 3 (including Reissues)	460	230	115
Multiple dependent claims	820	410	205

Total Claims		**Extra Claims**	**Fee ($)**		**Fee Paid ($)**
_____	- or HP =	_____	x _____	=	_____

HP = highest number of total claims paid for, if greater than 20.

Multiple Dependent Claims	
Fee ($)	**Fee Paid ($)**
_____	_____

Indep. Claims		**Extra Claims**	**Fee ($)**		**Fee Paid ($)**
_____	- or HP =	_____	x _____	=	_____

HP = highest number of independent claims paid for, if greater than 3.

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper (excluding electronically filed sequence or computer listings under 37 CFR 1.52(e)), the application size fee due is $400 ($200 for small entity) ($100 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets		**Extra Sheets**		**Number of each additional 50 or fraction thereof**	**Fee ($)**	**Fee Paid ($)**
_____	- 100 =	_____	/50 =	_____ (round **up** to a whole number)	x _____	= _____

4. OTHER FEE(S)

	Fees Paid ($)
Non-English specification, $130 fee (no small or micro entity discount)	
Non-electronic filing fee under 37 CFR 1.16(t) for a utility application, $400 fee ($200 small or micro entity)	
Other (e.g., late filing surcharge): 2251 Extension for response within first month	100.00

SUBMITTED BY

Signature	/Tani Chen/	Registration No. (Attorney/Agent)	52,728	Telephone	617.646.8000
Name (Print/Type)	Tani Chen, Sc.D.			Date	July 29, 2019

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: July 29, 2019 Electronic Signature for : /Eileen M. MacKenzie/

7433728.1

PTO/AIA/22 (03-13)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PETITION FOR EXTENSION OF TIME UNDER 37 CFR 1.136(a)	Docket Number (Optional) H0498.70509US02
Application Number 15/329,683-Conf. #7428	Filed January 27, 2017
For SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS	
Art Unit 1634	Examiner F. W. Lu

This is a request under the provisions of 37 CFR 1.136(a) to extend the period for filing a reply in the above-identified application.

The requested extension and fee are as follows (check time period desired and enter the appropriate fee below):

		Fee	Small Entity Fee	Micro Entity Fee	
[x]	One month (37 CFR 1.17(a)(1))	$200	$100	$50	$ 100.00
[ ]	Two months (37 CFR 1.17(a)(2))	$600	$300	$150	$
[ ]	Three months (37 CFR 1.17(a)(3))	$1,400	$700	$350	$
[ ]	Four months (37 CFR 1.17(a)(4))	$2,200	$1,100	$550	$
[ ]	Five months (37 CFR 1.17(a)(5))	$3,000	$1,500	$750	$

[x] Applicant asserts small entity status. See 37 CFR 1.27.

[ ] Applicant certifies micro entity status. See 37 CFR 1.29.
Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

[ ] A check in the amount of the fee is enclosed.

[x] Payment by credit card. ~~Form PTO-2038 is attached.~~

[ ] The Director has already been authorized to charge fees in this application to a Deposit Account.

[x] The Director is hereby authorized to charge any fees which may be required, or credit any overpayment, to Deposit Account Number 23/2825 .

[ ] Payment made via EFS-Web.

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

I am the

[ ] applicant.

[ ] attorney or agent of record. Registration number _____ .

[x] attorney or agent acting under 37 CFR 1.34. Registration number 52,728 .

/Tani Chen/ — Signature
July 29, 2019 — Date

Tani Chen, Sc.D. — Typed or printed name
617.646.8000 — Telephone Number

NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications. Submit multiple forms if more than one signature is required, see below*.

[ ] * Total of 1 forms are submitted.

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: July 29, 2019 Electronic Signature for : /Eileen M. MacKenzie/

7433729.1

DocCode – SCORE

SCORE Placeholder Sheet for IFW Content

Application Number: 15329683 Document Date: 07/29/2019

The presence of this form in the IFW record indicates that the following document type was received in electronic format on the date identified above. This content is stored in the SCORE database.

Since this was an electronic submission, there is no physical artifact folder, no artifact folder is recorded in PALM, and no paper documents or physical media exist. The TIFF images in the IFW record were created from the original documents that are stored in SCORE.

- Drawing

At the time of document entry (noted above):

- USPTO employees may access SCORE content via DAV or via the SCORE web page.
- External customers may access SCORE content via PAIR using the Supplemental Content tab.

Form Revision Date: March 1, 2019

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/329,683	01/27/2017	Xiaowei Zhuang	H0498.70509US02	7428

86110 7590 10/18/2019

Wolf, Greenfield & Sacks, P.C. (Harvard)
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER	
LU, FRANK WEI MIN	

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
10/18/2019	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

H0498_eOfficeAction@WolfGreenfield.com
Patents_eOfficeAction@WolfGreenfield.com

Office Action Summary	Application No. 15/329,683	Applicant(s) Zhuang et al.	
	Examiner FRANK W LU	Art Unit 1634	AIA (FITF) Status Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☑ Responsive to communication(s) filed on 7/29/2019.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☑ This action is **FINAL.** 2b) ☐ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5) ☑ Claim(s) 1,7,10,18,26,31,34,44,50,71,75,77-78,95-96,99,103,122,126 and 129 is/are pending in the application.

5a) Of the above claim(s) 95-96,99,103,122,126 and 129 is/are withdrawn from consideration.

6) ☐ Claim(s) _____ is/are allowed.

7) ☑ Claim(s) 1,7,10,18,26,31,34,44,50,71,75 and 77-78 is/are rejected.

8) ☐ Claim(s) _____ is/are objected to.

9) ☐ Claim(s) _____ are subject to restriction and/or election requirement

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to **PPHfeedback@uspto.gov.**

Application Papers

10)☐ The specification is objected to by the Examiner.

11)☑ The drawing(s) filed on 7/29/2019 is/are: a)☑ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☑ Notice of References Cited (PTO-892)

2) ☑ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date 7/29/2019.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date _____.

4) ☐ Other: _____.

DETAILED ACTION

Response to Amendment

1. Applicant's response to the office action filed on July 29, 2019 has been entered. The claims pending in this application are claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, 78, 95, 96, 99, 103, 122, 126, and 129 wherein claims 95, 96, 99, 103, 122, 126, and 129 have been withdrawn due to the restriction requirement mailed on September 24, 2018. Rejection and/or objection not reiterated from the previous office action are hereby withdrawn in view of applicant's amendment filed on July 29, 2019. Claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 will be examined.

Drawings

2. New Figures 20A to 20H filed on July 29, 2019 have been accepted by the office.

Claim Objections

3. Claim 1 is objected to because of the following informalities: (1) "each read sequence" in exposing step should be "each of the read sequences"; (2) "the codeword" in creating step should be "the codewords"; and (3) "the valid codewords" should be "the valid codeword".

4. Claim 31 is objected to because of the following informality: "read sequences" should be "the read sequences".

5. Claim 50 is objected to because of the following informality: "the identified nucleic acid target" should be "the nucleic acid targets".

Appropriate correction is required.

Claim Rejections - 35 USC § 101

6. 35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

7. Claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 are rejected under 35 U.S.C. 101 because the claimed invention is directed to a judicial exception (i.e., a law of nature, a natural phenomenon, or an abstract idea) without significantly more.

The processes of claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 are directed to a process of determining a nucleic acid abundance and/or spatial distribution within the sample. Claim 1 includes associating a plurality of nucleic acid targets in a sample with a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; exposing a sample suspected of containing nucleic acid to a plurality of nucleic acid probes, wherein at least some of the plurality of nucleic acid probes comprises a first portion comprising a target sequence and a second portion comprising one or more read sequences, wherein each of the read sequence represents a value of a position within the codewords; for each nucleic acid probe of the plurality of nucleic acid probes, determining binding of the nucleic acid probe within the sample; creating codewords corresponding to the binding of the plurality of nucleic acid probes within the sample, wherein the values of the digits of the codewords are based on the read sequences present on the nucleic acid probes; and for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, either discarding the codeword or applying error correction to the codeword to form a valid codeword, the valid codewords being a plurality of codewords assigned to the plurality of the nucleic acid targets; and determining a nucleic acid abundance and/or spatial distribution within the sample using the valid codewords corresponding

to the binding of the plurality of nucleic acid probes within the sample wherein associating step, creating step and the second determining step in independent claim 1 fall within the mental process grouping of abstract ideas. Dependent claim 44 comprises determining the centroid of the first signaling entity using an algorithm for determining non-overlapping single emitters and falls within the mental process grouping of abstract ideas. **Note that steps in a claim that recite the manipulation of information through a series of mental steps would be considered a single abstract idea for purposes of analysis** (see MPEP 2106.04).

This judicial exception, the abstract idea, is not integrated into a practical application because the additional elements, associating step, crating step and the second determining step in independent claim 1 and dependent claim 44 are mental steps that do not integrate the recited judicial exception into a practical application.

The claims do not include additional elements that are sufficient to amount to significantly more than the judicial exception because, since, in independent claim 68, exposing step only requires exposing a sample suspected of containing nucleic acid to a plurality of nucleic acid probes and the first determining step only requires determining binding of the nucleic acid probes within the sample wherein at least some of the plurality of nucleic acid probes comprises a first portion comprising a target sequence and a second portion comprising one or more read sequence and wherein each read sequence represents a value of a position within the codewords, exposing step and the first determining step are common, well-understood, routine and conventional methods used in the laboratory and are well known in the art (for example, see paragraphs [0007] and [0011] and claims 1 and 5 of US 2013/0190196 A1, published on July 25, 2013, and claim 47 of US Patent No. 7,807,447 B1, published on October 5, 2010). The conventional steps do not comprise an inventive concept that transforms the

claimed abstract ideas into a patent-eligible application of the abstract ideas. The claimed subject matters of claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 are processes directed to abstract ideas. When recited at this high level of generality, there is no meaningful limitation, such as a particular or unconventional machine or a transformation of a particular article, in this step that distinguishes it from well-understood, routine, and conventional data gathering activity engaged in by scientists prior to applicant's invention, and at the time the application was filed, *e.g.*, the routine and conventional techniques for performing exposing step and the first determining step of claim 1. Further, it is well established that the mere physical or tangible nature of additional elements such as exposing step and the first determining step of claim 1 do not automatically confer eligibility on a claim directed to an abstract idea (*see*, *e.g.*, *Alice Corp. v. CLS Bank Int'l*, 134 S. Ct. 2347, 2358-59 (2014)). Therefore, the claimed subject matters of claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 are without limitations that indicate that the claims are no different from the process claims directed to abstract ideas and are patent-ineligible.

Claim Rejections - 35 USC § 112

8. The following is a quotation of the first paragraph of 35 U.S.C. 112(a):

> (a) IN GENERAL.—The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor or joint inventor of carrying out the invention.

The following is a quotation of the first paragraph of pre-AIA 35 U.S.C. 112:

> The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor of carrying out his invention.

9. Scope of Enablement

Claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 are rejected under 35 U.S.C. 112(a) or 35 U.S.C. 112 (pre-AIA), first paragraph, because the specification, while being enabling for exposing a sample suspected of containing nucleic acid to a plurality of nucleic acid probes, does not reasonably provide enablement for performing the methods recited in claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78. The specification does not enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to use the invention commensurate in scope with these claims.

In *In re Wands*, 858 F.2d 731,737, 8 USPQ2d 1400, 1404 (Fed. Cir. 1988) the court considered the issue of enablement in molecular biology. The Court summarized eight factors to be considered in a determination of "undue experimentation". These factors include: (a) the quantity of experimentation necessary; (b) the amount of direction or guidance presented; (c) the presence or absence of working examples; (d) the nature of the invention; (e) the state of the prior art; (f) the relative skill of those in the art; (g) the predictability of the art; and (h) the breadth of the claims. The Court also stated that although the level of skill in molecular biology is high, results of experiments in molecular biology are unpredictable.

To begin, there is no direction or guidance to show that the methods recited in claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 can be performed. While the relative skill in the art is very high (the Ph.D. degree with laboratory experience), there is no predictability whether the methods recited in claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 can be performed.

First, since claim 1 does not indicate how to correlate the binding of the plurality of nucleic acid probes within the sample with codewords, it is unclear how codewords

corresponding to the binding of the plurality of nucleic acid probes can be created. Furthermore, although claim 1 requires the valid codewords being a plurality of codewords assigned to the plurality of the nucleic acid targets, it is unclear that, in which situation, at least some of the codeword matches with a valid codeword and in which situation, at least some of the codeword does not match with a valid codeword. In addition, since claim 1 does not indicate how a nucleic acid abundance and/or spatial distribution within the sample is correlate with the valid codewords corresponding to the binding of the plurality of nucleic acid probes within the sample, it is unclear how a nucleic acid abundance and/or spatial distribution within the sample can be determined using the valid codewords corresponding to the binding of the plurality of nucleic acid probes within the sample as recited in claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 and how an error-correcting code can be defined when the plurality of read sequences are distributed on the plurality of nucleic acid probes as recited in claim 18.

Second, since claim 26 does not indicate what is the relationship between the first secondary probe and the nucleic acid probes in claim 1, and claim 31 does not indicate what is the relationship between the second secondary probe and the nucleic acid probes in claim 1 and how the first signaling entity is different from the second signaling entity, the first secondary probe and/or the second secondary probe, and the nucleic acid probes are different nucleic acid probes, it is unclear how binding of the nucleic acid probes can be determined by the first signaling entity within the sample as recited in claim 26 and how binding of the nucleic acid probes can be determined by the second signaling entity within the sample as recited in claim 31.

Third, since claim 44 does not require that the first signaling entity is a single emitter which can emit a nonoverlapping signal, it is unclear how the centroid of the first signaling entity can be determined using an algorithm for determining nonoverlapping single emitters.

Fourth, since claim 78 does not indicate what kind of microscope is used to determine binding of the nucleic acid probes within the sample, it is unclear how the binding of the nucleic acid probes can be determined by imaging at least a portion of the sample using any kind of microscope such as low resolution light microscope.

In view of above discussions, the skilled artisan will have no way to predict the experimental results. Accordingly, it is concluded that undue experimentation is required to make the invention as it is claimed. These undue experimentation at least includes to test whether the methods recited in claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 can be performed.

10. The following is a quotation of 35 U.S.C. 112(b):

> (b) CONCLUSION.—The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the inventor or a joint inventor regards as the invention.

> The following is a quotation of 35 U.S.C. 112 (pre-AIA), second paragraph:
> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

11. Claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 are rejected under 35 U.S.C. 112(b) or 35 U.S.C. 112 (pre-AIA), second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which the inventor or a joint inventor, or for pre-AIA the applicant regards as the invention.

12. Claim 1 is rejected as vague and indefinite because it is unclear what is the relationship between a sample in associating step and a sample suspected of containing nucleic acid in exposing step. Please clarify.

13. Claim 26 or 31 is rejected as vague and indefinite because it is unclear that unbound probe comprises what probe. Please clarify.

14. The following is a quotation of 35 U.S.C. 112(d):

> (d) REFERENCE IN DEPENDENT FORMS.—Subject to subsection (e), a claim in dependent form shall contain a reference to a claim previously set forth and then specify a further limitation of the subject matter claimed. A claim in dependent form shall be construed to incorporate by reference all the limitations of the claim to which it refers.

The following is a quotation of pre-AIA 35 U.S.C. 112, fourth paragraph:

> Subject to the following paragraph [i.e., the fifth paragraph of pre-AIA 35 U.S.C. 112], a claim in dependent form shall contain a reference to a claim previously set forth and then specify a further limitation of the subject matter claimed. A claim in dependent form shall be construed to incorporate by reference all the limitations of the claim to which it refers.

15. Claim 10 is rejected under 35 U.S.C. 112(d) or pre-AIA 35 U.S.C. 112, 4th paragraph, as being of improper dependent form for failing to further limit the subject matter of the claim upon which it depends, or for failing to include all the limitations of the claim upon which it depends. Since claim 1 has all limitations of claim 10, applicant may cancel the claim(s), amend the claim(s) to place the claim(s) in proper dependent form, rewrite the claim(s) in independent form, or present a sufficient showing that the dependent claim(s) complies with the statutory requirements.

Response to Arguments

16. Applicant's arguments with respect to claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 have been considered but are moot because the arguments do not apply to the current rejection.

Conclusion

17. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

18. No claim is allowed.

19. Papers related to this application may be submitted to Group 1600 by facsimile transmission. Papers should be faxed to Group 1600 via the PTO Fax Center. The faxing of such papers must conform with the notices published in the Official Gazette, 1096 OG 30 (November 15, 1988), 1156 OG 61 (November 16, 1993), and 1157 OG 94 Fax Center number is (571)273-8300.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Frank Lu, Ph.D., whose telephone number is (571)272-0746. The examiner can normally be reached on Monday-Friday from 9 A.M. to 5 P.M.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Dave Nguyen, can be reached on (571)272-0731.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR

system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/FRANK W LU/
Primary Examiner, Art Unit 1634
October 11, 2019

Notice of References Cited	Application/Control No. 15/329,683	Applicant(s)/Patent Under Reexamination Zhuang et al.	
	Examiner FRANK W LU	Art Unit 1634	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	CPC Classification	US Classification
*	A	US-20130190196-A1	07-2013	Onderdonk; Andrew B.	C12Q1/6888	506/9
*	B	US-7807447-B1	10-2010	Shoemaker; Daniel D.	C12Q1/6809	435/287.1
	C					
	D					
	E					
	F					
	G					
	H					
	I					
	J					
	K					
	L					
	M					

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	CPC Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15/329,683	Zhuang et al.
	Examiner	Art Unit
	FRANK W LU	1634

CPC - Searched*		
Symbol	Date	Examiner
Update	10/11/2019	/FL/

CPC Combination Sets - Searched*		
Symbol	Date	Examiner
Update	10/11/2019	/FL/

US Classification - Searched*			
Class	Subclass	Date	Examiner
Update		10/11/2019	/FL/

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

Search Notes		
Search Notes	Date	Examiner
Update	10/11/2019	/FL/

Interference Search			
US Class/CPC Symbol	US Subclass/CPC Group	Date	Examiner

/FRANK W LU/ Primary Examiner, Art Unit 1634	October 11, 2019

PART II: Information Cited

The undersigned hereby makes of record in the above-identified application the information listed on the attached form PTO-1449 (modified PTO/SB/08). The order of presentation of the references should not be construed as an indication of the importance of the references.

The undersigned hereby makes the following additional information of record in the above-identified application.

The undersigned would like to bring to the Examiner's attention the following co-pending applications that may contain subject matter related to this application:

Serial No.	Filing Date	Inventor(s)	Docket No.
*16/253,919	1/22/2019	Zhuang et al.	H0498.70510US02
*16/347,874	5/7/2019	Zhuang et al.	H0498.70630US01
*16/348,071	5/7/2019	Zhuang et al.	H0498.70645US01

*A copy of this reference is not provided as the Office has waived the requirement under 37 C.F.R. § 1.98(a)(2)(iii) for submitting a copy of a cited U.S. patent application if it is scanned to the Image File Wrapper system and is available on Private PAIR.

The undersigned would like to bring to the Examiner's attention the enclosed search report or other communication from a corresponding or related International or Foreign National Application:

Serial No.	Date of Mailing	Type(s) of Communication	Docket No.
CN 201580052678.7	**01-16-2019**	**Chinese Office Action**	**H0498.70509CN00**
PCT/US17/60570	**05-23-2019**	**International Preliminary Report on Patentability**	**H0498.70630WO00**
PCT/US17/60558	**05-23-2019**	**International Preliminary Report on Patentability**	**H0498.70645WO00**

PART III: Remarks

Documents cited anywhere in the Information Disclosure Statement are enclosed unless otherwise indicated. It is respectfully requested that:

10/14/2019

/FRANK W LU/

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

FORM PTO-1449/A and B (modified PTO/SB/08)				APPLICATION NO.: 15/329,683	ATTY. DOCKET NO.: H0498.70509US02
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				FILING DATE: January 27, 2017	CONFIRMATION NO.: 7428
				FIRST NAMED INVENTOR: Xiaowei Zhuang	
Sheet	1	of	1	GROUP ART UNIT: 1634	EXAMINER: Frank Wei Min Lu

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document		Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		Number	Kind Code		
		7,776,613		Zhuang et al.	08-17-2010
		10,240,146	A1	Zhuang et al.	03-26-2019
		2016-0370295		Zhuang et al.	12-22-2016

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document			Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)
		Office/ Country	Number	Kind Code			
		WO	2008/091296	A2	President and Fellows of Harvard College	07-31-2008	
		WO	2016/018963	A9	President and Fellows of Harvard College	02-04-2016	

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		Chinese Office Action mailed January 16, 2019 for Application No. 201580052678.7 (H0498.70509CN00).	Y - Full
		International Preliminary report on Patentability mailed May 23, 2019 for Application No. PCT/US17/60570 (H0498.70630WO00).	
		International Preliminary report on Patentability mailed May 23, 2019 for Application No. PCT/US17/60558 (H0498.70645WO00).	
		LEVSKY et al., Single-cell gene expression profiling. Science. 2002 Aug 2; 297:836-840.	
		PLAYER et al., Single-copy gene detection using branched DNA (bDNA) in situ hybridization. The Journal of Histochemistry & Cytochemistry. 2001; 49(5): 603-611.	
		ZHEN et al., Poly-fish: A technique of repeated hybridizations that improves cytogenetic analysis of fetal cells in maternal blood. Prenatal Diagnosis. 1998 Mar 21; 18: 1181-1185.	

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

EXAMINER:	DATE CONSIDERED:
/FRANK W LU/	10/14/2019

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

FORM PTO-1449/A and B (modified PTO/SB/08)				APPLICATION NO.: 15/329,683	ATTY. DOCKET NO.: H0498.70509US02
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				FILING DATE: January 27, 2017	CONFIRMATION NO.: 7428
				FIRST NAMED INVENTOR: Xiaowei Zhuang	
Sheet	1	of	1	GROUP ART UNIT: 1634	EXAMINER: F. W. Lu

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document		Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		Number	Kind Code		

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document			Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)
		Office/ Country	Number	Kind Code			

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		European Office Action mailed October 16, 2019 for Application No. 15827358.1 (H0498.70509EP00).	

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

EXAMINER:	DATE CONSIDERED:

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

7681167.1

Electronic Patent Application Fee Transmittal

Application Number:	15329683
Filing Date:	27-Jan-2017
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Filer:	Tani Chen/Catherine Borgnis
Attorney Docket Number:	H0498.70509US02

Filed as Small Entity

Filing Fees for U.S. National Stage under 35 USC 371

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
SUBMISSION- INFORMATION DISCLOSURE STMT	2806	1	120	120
		Total in USD ($)		**120**

Electronic Acknowledgement Receipt

EFS ID:	37615470
Application Number:	15329683
International Application Number:	
Confirmation Number:	7428
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Customer Number:	86110
Filer:	Tani Chen/Catherine Borgnis
Filer Authorized By:	Tani Chen
Attorney Docket Number:	H0498.70509US02
Receipt Date:	31-OCT-2019
Filing Date:	27-JAN-2017
Time Stamp:	10:30:24
Application Type:	U.S. National Stage under 35 USC 371

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$120
RAM confirmation Number	E20190UA32240146
Deposit Account	232825
Authorized User	Catherine Borgnis
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows: 37 CFR 1.17 (Patent application and reexamination processing fees)	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	H049870509US02-TRN-TC.pdf	25971 513aa4473dec9806977377 64f923946ad72bb3ee	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	H049870509US02-FEE-TC.pdf	32725 0d661126af12ae7366eb170d3e5857e580973289	no	1
Warnings:					
Information:					
3	Transmittal Letter	H049870509US02-IDSBODY-TC.pdf	32825 5fe01bb23a7bb23668bb56ccd52479abf231a8f3	no	4
Warnings:					
Information:					
4	Information Disclosure Statement (IDS) Form (SB08)	H049870509US02-IDS1449-TC.pdf	21077 81b29529b530f01554735b2ef5ec02447e6ca998	no	1
Warnings:					
Information:					
This is not an USPTO supplied IDS fillable form					
5	Other Reference-Patent/App/Search documents	H049870509EP00_EP_OA_mailed_10-16-2019.pdf	131313 a43c7030ce60e6991644d1d94140a1e2ac9a3121	no	4
Warnings:					
Information:					
6	Fee Worksheet (SB06)	fee-info.pdf	30672 754e7ef91c646cfca28b7dae371ca69e5e6c739e	no	2
Warnings:					

Information:	
Total Files Size (in bytes):	274583

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/329,683-Conf. #7428
Filing Date	January 27, 2017
First Named Inventor	Xiaowei Zhuang
Art Unit	1634
Examiner Name	F. W. Lu
Attorney Docket Number	H0498.70509US02

ENCLOSURES (*Check all that apply*)

- [x] Fee Transmittal Form
  - [ ] Fee Attached
- [ ] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [x] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53

- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD

- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) (please Identify below):

Copy of Cited Reference
Form PTO-1449

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Tani Chen/		
Printed name	Tani Chen, Sc.D.		
Date	October 31, 2019	Reg. No.	52,728

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: ___October 31, 2019_____ Electronic Signature for Catherine E. Borgnis: _/Catherine E. Borgnis/_

7681166.1

FEE TRANSMITTAL

Complete if known	
Application Number	15/329,683-Conf. #7428
Filing Date	January 27, 2017
First Named Inventor	Xiaowei Zhuang
Examiner Name	F. W. Lu
Art Unit	1634
Practitioner Docket No.	H0498.70509US02

[X] Applicant asserts small entity status. See 37 CFR 1.27.

[ ] Applicant certifies micro entity status. See 37 CFR 1.29. Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

TOTAL AMOUNT OF PAYMENT ($)120.00

METHOD OF PAYMENT (check all that apply)

[ ] Check [X] Credit Card [ ] Money Order [ ] None [ ] Other (please identify): _____

[X] Deposit Account Deposit Account Number: 23/2825 Deposit Account Name: Wolf, Greenfield & Sacks, P.C.

For the above-identified deposit account, the Director is hereby authorized to (check all that apply):

[X] Charge fee(s) indicated below

[ ] Charge fee(s) indicated below, **except for the filing fee**

[X] Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17, **except for any excess claims fees or multiple dependent claim fee**

[ ] Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17

[X] Credit any overpayment of fee(s)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES (U = undiscounted fee; S = small entity fee; M = micro entity fee)

	FILING FEES			SEARCH FEES			EXAMINATION FEES			
Application Type	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**Fees Paid ($)**
Utility	300	150*	75	660	330	165	760	380	190	
Design	200	100	50	160	80	40	600	300	150	
Plant	200	100	50	420	210	105	620	310	155	
Reissue	300	150	75	660	330	165	2,200	1,100	550	
Provisional	280	140	70	0	0	0	0	0	0	

* The $150 small entity status filing fee for a utility application is further reduced to $75 for a small entity status applicant who files the application via EFS-Web.

2. EXCESS CLAIM FEES

Fee Description	**Undiscounted Fee ($)**	**Small Entity Fee ($)**	**Micro Entity Fee ($)**
Each claim over 20 (including Reissues)	100	50	25
Each independent claim over 3 (including Reissues)	460	230	115
Multiple dependent claims	820	410	205

Total Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**
_____ - 20 or HP =	_____ x	_____ =	_____

HP = highest number of total claims paid for, if greater than 20.

Indep. Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**
_____ - 3 or HP =	_____ x	_____ =	_____

HP = highest number of independent claims paid for, if greater than 3.

Multiple Dependent Claims	
Fee ($)	**Fee Paid ($)**
_____	_____

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper (excluding electronically filed sequence or computer listings under 37 CFR 1.52(e)), the application size fee due is $400 ($200 for small entity) ($100 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets	**Extra Sheets**	**Number of each additional 50 or fraction thereof**	**Fee ($)**	**Fee Paid ($)**
_____ - 100 =	_____ /50 =	_____ (round **up** to a whole number) x	_____ =	_____

4. OTHER FEE(S)

	Fees Paid ($)
Non-English specification, $130 fee (no small or micro entity discount)	
Non-electronic filing fee under 37 CFR 1.16(t) for a utility application, $400 fee ($200 small or micro entity)	
Other (e.g., late filing surcharge): 2806 Submission of an Information Disclosure Statement	120.00

SUBMITTED BY

Signature	/Tani Chen/	Registration No. (Attorney/Agent)	52,728	Telephone	617.646.8000
Name (Print/Type)	Tani Chen, Sc.D.			Date	October 31, 2019

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: _____October 31, 2019_____ Electronic Signature for Catherine E. Borgnis: _/Catherine E. Borgnis/_

7681165.1

DOCKET NO.: H0498.70509US02

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	Xiaowei Zhuang
Application No.:	15/329,683
Confirmation No.:	7428
Filed:	January 27, 2017
For:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
Examiner:	F. W. Lu
Art Unit:	1634

Certificate of Electronic Filing under 37 C.F.R. § 1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: ____October 31, 2019_____ Electronic Signature for Catherine E. Borgnis: _/Catherine E. Borgnis/_

MAIL STOP AMENDMENT
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

STATEMENT FILED PURSUANT TO THE DUTY OF DISCLOSURE UNDER 37 C.F.R. §§ 1.56, 1.97 AND 1.98

Sir:

Pursuant to the duty of disclosure under 37 C.F.R. §§ 1.56, 1.97 and 1.98, the undersigned requests consideration of this Information Disclosure Statement.

PART I: Compliance with 37 C.F.R. § 1.97

This Information Disclosure Statement has been filed after the mailing date of either a Final Action under 37 C.F.R. § 1.113 or a Notice of Allowance under 37 C.F.R. § 1.311 and is being filed on or before payment of an Issue Fee.

The undersigned hereby states, as specified in 37 C.F.R. § 1.97(e), that each item of information contained in this Information Disclosure Statement was first cited in any communication from a foreign patent office in a counterpart for this application not more than three months prior to the filing of this Statement.

Please charge our Credit Card in the amount of $120.00 covering the fee set forth in 37 C.F.R. § 1.17(p).

PART II: Compliance with 37 C.F.R. § 1.704(d)

Each item of information contained in this Information Disclosure Statement

(i) was first cited in any communication from a patent office in a counterpart foreign or international application or from the United States Patent and Trademark Office, and this communication was not received by any individual designated in 37 C.F.R. § 1.56(c) more than thirty (30) days prior to the filing of the Information Disclosure Statement; or

(ii) is a communication that was issued by a patent office in a counterpart foreign or international application or by the United States Patent and Trademark Office, and this communication was not received by any individual designated in 37 C.F.R. § 1.56(c) more than thirty (30) days prior to the filing of the Information Disclosure Statement.

PART III: Information Cited

The undersigned hereby makes of record in the above-identified application the information listed on the attached form PTO-1449 (modified PTO/SB/08). The order of presentation of the references should not be construed as an indication of the importance of the references.

The undersigned hereby makes the following additional information of record in the above-identified application.

The undersigned would like to bring to the Examiner's attention the enclosed search report or other communication from a corresponding or related International or Foreign National Application:

Serial No.	Date of Mailing	Type(s) of Communication	Docket No.
EP 15827358.1	**10-16-2019**	**European Office Action**	**H0498.70509EP00**

PART IV: Remarks

Documents cited anywhere in the Information Disclosure Statement are enclosed unless otherwise indicated. It is respectfully requested that:

1. The Examiner consider completely the cited information, along with any other information, in reaching a determination concerning the patentability of the present claims;

2. The enclosed form PTO-1449 (modified PTO/SB/08) be signed by the Examiner to evidence that the cited information has been fully considered by the United States Patent and Trademark Office during the examination of this application;

3. The citations for the information be printed on any patent which issues from this application.

By submitting this Information Disclosure Statement, the undersigned makes no representation that a search has been performed, of the extent of any search performed, or that more relevant information does not exist.

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, material to patentability as defined in 37 C.F.R. § 1.56(b).

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, in fact, prior art as defined by 35 U.S.C. § 102.

Notwithstanding any statements by the undersigned, the Examiner is urged to form his or her own conclusion regarding the relevance of the cited information.

An early and favorable action is hereby requested.

The Director is hereby authorized to charge any deficiency or credit any overpayment in the fees occasioned by the filing of this Information Disclosure Statement to our Deposit Account No. 23/2825 under Docket No. H0498.70509US02 from which the undersigned is authorized to draw.

Respectfully submitted,

Electronic signature: /Tani Chen/
Tani Chen, Sc.D.
Registration No.: 52,728
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

Docket No.: H0498.70509US02
Date: October 31, 2019

FORM PTO-1449/A and B (modified PTO/SB/08)				APPLICATION NO.: 15/329,683	ATTY. DOCKET NO.: H0498.70509US02
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				FILING DATE: January 27, 2017	CONFIRMATION NO.: 7428
				FIRST NAMED INVENTOR: Xiaowei Zhuang	
Sheet	1	of	1	GROUP ART UNIT: 1634	EXAMINER: F. W. Lu

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document		Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		Number	Kind Code		

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document			Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)
		Office/ Country	Number	Kind Code			

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		Chinese Office Action dated November 14, 2019 for Application No. 201580052678.7 (H0498.70509CN00).	Y - Full

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

EXAMINER:	DATE CONSIDERED:

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

7786753.1

Electronic Patent Application Fee Transmittal

Application Number:	15329683
Filing Date:	27-Jan-2017
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Filer:	Tani Chen/Catherine Borgnis
Attorney Docket Number:	H0498.70509US02

Filed as Small Entity

Filing Fees for U.S. National Stage under 35 USC 371

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
SUBMISSION- INFORMATION DISCLOSURE STMT	2806	1	120	120
		Total in USD ($)		**120**

Electronic Acknowledgement Receipt

EFS ID:	37998403
Application Number:	15329683
International Application Number:	
Confirmation Number:	7428
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Customer Number:	86110
Filer:	Tani Chen/Catherine Borgnis
Filer Authorized By:	Tani Chen
Attorney Docket Number:	H0498.70509US02
Receipt Date:	11-DEC-2019
Filing Date:	27-JAN-2017
Time Stamp:	14:30:49
Application Type:	U.S. National Stage under 35 USC 371

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$120
RAM confirmation Number	E2019BAE31271490
Deposit Account	232825
Authorized User	Catherine Borgnis
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows: 37 CFR 1.17 (Patent application and reexamination processing fees)	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	H049870509US02-TRN-TC.pdf	26148 0df3fa48ffe578d1e6e11292a97fd137e1572e5e	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	H049870509US02-FEE-TC.pdf	32711 1905f7f21e4d1f5fd5b19ca4081bde148fe293dc	no	1
Warnings:					
Information:					
3	Transmittal Letter	H049870509US02-IDSBODY-TC.pdf	32477 052fdb5e437e01dd7a9bff6054171aa33b16622b	no	4
Warnings:					
Information:					
4	Information Disclosure Statement (IDS) Form (SB08)	H049870509US02-IDS1449-TC.pdf	21342 6f9ed3d6b17123d902a7d5ac21d7bd77b163f2a1	no	1
Warnings:					
Information:					
This is not an USPTO supplied IDS fillable form					
5	Other Reference-Patent/App/Search documents	H049870509CN00_OA_mailed_11-14-2019_QC.pdf	343628 e0b51a67777f8520b037e3baa1b23ae65f9c4e23	no	6
Warnings:					
Information:					
6	Fee Worksheet (SB06)	fee-info.pdf	30672 3234a3737c9e43ab426c38afba6c39f31730a882	no	2
Warnings:					

Information:	
Total Files Size (in bytes):	486978

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/329,683-Conf. #7428
Filing Date	January 27, 2017
First Named Inventor	Xiaowei Zhuang
Art Unit	1634
Examiner Name	F. W. Lu
Attorney Docket Number	H0498.70509US02

ENCLOSURES (*Check all that apply*)

- [x] Fee Transmittal Form
  - [ ] Fee Attached
- [ ] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [x] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53

- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD

- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) (please Identify below):

Copy of Cited Reference
Form PTO/1449

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Tani Chen/		
Printed name	Tani Chen, Sc.D.		
Date	December 11, 2019	Reg. No.	52,728

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: ___December 11, 2019_____ Electronic Signature for Catherine E. Borgnis: _/Catherine E. Borgnis/_

7786752.1

FEE TRANSMITTAL

Complete if known	
Application Number	15/329,683-Conf. #7428
Filing Date	January 27, 2017
First Named Inventor	Xiaowei Zhuang
Examiner Name	F. W. Lu
Art Unit	1634
Practitioner Docket No.	H0498.70509US02

[X] Applicant asserts small entity status. See 37 CFR 1.27.

[ ] Applicant certifies micro entity status. See 37 CFR 1.29. Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

TOTAL AMOUNT OF PAYMENT ($)120.00

METHOD OF PAYMENT (check all that apply)

[ ] Check [X] Credit Card [ ] Money Order [ ] None [ ] Other (please identify): _____

[X] Deposit Account Deposit Account Number: 23/2825 Deposit Account Name: Wolf, Greenfield & Sacks, P.C.

For the above-identified deposit account, the Director is hereby authorized to (check all that apply):

[X] Charge fee(s) indicated below

[ ] Charge fee(s) indicated below, **except for the filing fee**

[X] Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17, **except for any excess claims fees or multiple dependent claim fee**

[ ] Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17

[X] Credit any overpayment of fee(s)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES (U = undiscounted fee; S = small entity fee; M = micro entity fee)

	FILING FEES			SEARCH FEES			EXAMINATION FEES			
Application Type	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**Fees Paid ($)**
Utility	300	150*	75	660	330	165	760	380	190	
Design	200	100	50	160	80	40	600	300	150	
Plant	200	100	50	420	210	105	620	310	155	
Reissue	300	150	75	660	330	165	2,200	1,100	550	
Provisional	280	140	70	0	0	0	0	0	0	

* The $150 small entity status filing fee for a utility application is further reduced to $75 for a small entity status applicant who files the application via EFS-Web.

2. EXCESS CLAIM FEES

Fee Description	**Undiscounted Fee ($)**	**Small Entity Fee ($)**	**Micro Entity Fee ($)**
Each claim over 20 (including Reissues)	100	50	25
Each independent claim over 3 (including Reissues)	460	230	115
Multiple dependent claims	820	410	205

Total Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**
_____ - 20 or HP =	_____ x	_____ =	_____

HP = highest number of total claims paid for, if greater than 20.

Indep. Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**
_____ - 3 or HP =	_____ x	_____ =	_____

HP = highest number of independent claims paid for, if greater than 3.

Multiple Dependent Claims	
Fee ($)	**Fee Paid ($)**

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper (excluding electronically filed sequence or computer listings under 37 CFR 1.52(e)), the application size fee due is $400 ($200 for small entity) ($100 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets	**Extra Sheets**	**Number of each additional 50 or fraction thereof**	**Fee ($)**	**Fee Paid ($)**
_____ - 100 =	_____ /50 =	_____ (round **up** to a whole number) x	_____ =	_____

4. OTHER FEE(S)

	Fees Paid ($)
Non-English specification, $130 fee (no small or micro entity discount)	
Non-electronic filing fee under 37 CFR 1.16(t) for a utility application, $400 fee ($200 small or micro entity)	
Other (e.g., late filing surcharge): 2806 Submission of an Information Disclosure Statement	120.00

SUBMITTED BY

Signature	/Tani Chen/	Registration No. (Attorney/Agent)	52,728	Telephone	617.646.8000
Name (Print/Type)	Tani Chen, Sc.D.			Date	December 11, 2019

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: ____December 11, 2019_____ Electronic Signature for Catherine E. Borgnis: _/Catherine E. Borgnis/_

7786751.1

DOCKET NO.: H0498.70509US02

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	Xiaowei Zhuang
Application No.:	15/329,683
Confirmation No.:	7428
Filed:	January 27, 2017
For:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
Examiner:	F. W. Lu
Art Unit:	1634

Certificate of Electronic Filing under 37 C.F.R. § 1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: __December 11, 2019____ Electronic Signature for Catherine E. Borgnis: _/Catherine E. Borgnis/_

MAIL STOP AMENDMENT
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

STATEMENT FILED PURSUANT TO THE DUTY OF DISCLOSURE UNDER 37 C.F.R. §§ 1.56, 1.97 AND 1.98

Sir:

Pursuant to the duty of disclosure under 37 C.F.R. §§ 1.56, 1.97 and 1.98, the undersigned requests consideration of this Information Disclosure Statement.

PART I: Compliance with 37 C.F.R. § 1.97

This Information Disclosure Statement has been filed after the mailing date of either a Final Action under 37 C.F.R. § 1.113 or a Notice of Allowance under 37 C.F.R. § 1.311 and is being filed on or before payment of an Issue Fee.

The undersigned hereby states, as specified in 37 C.F.R. § 1.97(e), that each item of information contained in this Information Disclosure Statement was first cited in any communication from a foreign patent office in a counterpart for this application not more than three months prior to the filing of this Statement.

Please charge our Credit Card in the amount of $120.00 covering the fee set forth in 37 C.F.R. § 1.17(p).

PART II: Compliance with 37 C.F.R. § 1.704(d)

Each item of information contained in this Information Disclosure Statement

(i) was first cited in any communication from a patent office in a counterpart foreign or international application or from the United States Patent and Trademark Office, and this communication was not received by any individual designated in 37 C.F.R. § 1.56(c) more than thirty (30) days prior to the filing of the Information Disclosure Statement; or

(ii) is a communication that was issued by a patent office in a counterpart foreign or international application or by the United States Patent and Trademark Office, and this communication was not received by any individual designated in 37 C.F.R. § 1.56(c) more than thirty (30) days prior to the filing of the Information Disclosure Statement.

PART III: Information Cited

The undersigned hereby makes of record in the above-identified application the information listed on the attached form PTO-1449 (modified PTO/SB/08). The order of presentation of the references should not be construed as an indication of the importance of the references.

The undersigned hereby makes the following additional information of record in the above-identified application.

The undersigned would like to bring to the Examiner's attention the enclosed search report or other communication from a corresponding or related International or Foreign National Application:

Serial No.	Date of Mailing	Type(s) of Communication	Docket No.
CN 201580052678.7	**11-14-2019**	**Chinese Office Action**	**H0498.70509CN00**

PART IV: Remarks

Documents cited anywhere in the Information Disclosure Statement are enclosed unless otherwise indicated. It is respectfully requested that:

1. The Examiner consider completely the cited information, along with any other information, in reaching a determination concerning the patentability of the present claims;

2. The enclosed form PTO-1449 (modified PTO/SB/08) be signed by the Examiner to evidence that the cited information has been fully considered by the United States Patent and Trademark Office during the examination of this application;

3. The citations for the information be printed on any patent which issues from this application.

By submitting this Information Disclosure Statement, the undersigned makes no representation that a search has been performed, of the extent of any search performed, or that more relevant information does not exist.

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, material to patentability as defined in 37 C.F.R. § 1.56(b).

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, in fact, prior art as defined by 35 U.S.C. § 102.

Notwithstanding any statements by the undersigned, the Examiner is urged to form his or her own conclusion regarding the relevance of the cited information.

An early and favorable action is hereby requested.

The Director is hereby authorized to charge any deficiency or credit any overpayment in the fees occasioned by the filing of this Information Disclosure Statement to our Deposit Account No. 23/2825 under Docket No. H0498.70509US02 from which the undersigned is authorized to draw.

Respectfully submitted,

Electronic signature: /Tani Chen/
Tani Chen, Sc.D.
Registration No.: 52,728
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

Docket No.: H0498.70509US02
Date: December 11, 2019

Request for Continued Examination (RCE) Transmittal		
Address to: Mail Stop RCE Commissioner for Patents P.O. Box 1450 Alexandria, VA 22313-1450	Application Number	15/329,683-Conf. #7428
	Filing Date	January 27, 2017
	First Named Inventor	Xiaowei Zhuang
	Art Unit	1634
	Examiner Name	F. W. Lu
	Attorney Docket Number	H0498.70509US02

This is a Request for Continued Examination (RCE) under 37 CFR 1.114 of the above-identified application.

Request for Continued Examination (RCE) practice under 37 CFR 1.114 does not apply to any utility or plant application filed prior to June 8, 1995, to any international application that does not comply with the requirements of 35 U.S.C 371, or to any design application. See Instruction Sheet for RCEs (not to be submitted to the USPTO on page 2.)

1. Submission required under 37 CFR 1.114 Note: If the RCE is proper, any previously filed unentered amendments and amendments enclosed with the RCE will be entered in the order in which they were filed unless applicant instructs otherwise. If applicant does not wish to have any previously filed unentered amendment(s) entered, applicant must request non-entry of such amendment(s).
   - a. [ ] Previously submitted. If a final Office action is outstanding, any amendments filed after the final Office action may be considered as a submission even if this box is not checked.
     - i. [ ] Consider the arguments in the Appeal Brief or Reply Brief previously filed on _____
     - ii. [ ] Other _____
   - b. [x] Enclosed
     - i. [x] Amendment/Reply
     - ii. [ ] Affidavit(s)/ Declaration(s)
     - iii. [x] Information Disclosure Statement (IDS)
     - iv. [ ] Other _____
2. Miscellaneous
   - a. [ ] Suspension of action on the above-identified application is requested under 37 CFR 1.103(c) for a period of _____ months. (Period of suspension shall not exceed 3 months; Fee under 37 CFR 1.17(i) required)
   - b. [ ] Other _____
3. Fees The RCE fee under 37 CFR 1.17(e) is required by 37 CFR 1.114 when the RCE is filed.
   - a. [x] The Director is hereby authorized to charge the following fees, any underpayment of fees, or credit any overpayments, to Deposit Account No. 23/2825 .
     - i. [x] RCE fee required under 37 CFR 1.17(e)
     - ii. [x] Extension of time fee (37 CFR 1.136 and 1.17)
     - iii. [ ] Other _____
   - b. [ ] Check in the amount of $ _____ enclosed
   - c. [x] Payment by credit card ~~(Form PTO-2038 enclosed)~~

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT REQUIRED			
Signature	/Tani Chen/	Date	April 17, 2020
Name (Print/Type)	Tani Chen, Sc.D.	Registration No.	52,728

CERTIFICATE OF MAILING OR TRANSMISSION			
I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Mail Stop RCE, Commissioner for Patents, P. O. Box 1450, Alexandria, VA 22313-1450 or facsimile transmitted to the U.S. Patent and Trademark Office on the date shown below.			
Signature			
Name (Print/Type)		Date	

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: April 17, 2020 Electronic Signature for Sarah Carletti: /Sarah Carletti/

8077149.1

PTO/AIA/22 (03-13)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PETITION FOR EXTENSION OF TIME UNDER 37 CFR 1.136(a)	Docket Number (Optional) H0498.70509US02
Application Number 15/329,683-Conf. #7428	Filed January 27, 2017
For SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS	
Art Unit 1634	Examiner F. W. Lu

This is a request under the provisions of 37 CFR 1.136(a) to extend the period for filing a reply in the above-identified application.

The requested extension and fee are as follows (check time period desired and enter the appropriate fee below):

	Fee	Small Entity Fee	Micro Entity Fee	
☐ One month (37 CFR 1.17(a)(1))	$200	$100	$50	$
☐ Two months (37 CFR 1.17(a)(2))	$600	$300	$150	$
☒ Three months (37 CFR 1.17(a)(3))	$1,400	$700	$350	$ 700.00
☐ Four months (37 CFR 1.17(a)(4))	$2,200	$1,100	$550	$
☐ Five months (37 CFR 1.17(a)(5))	$3,000	$1,500	$750	$

☒ Applicant asserts small entity status. See 37 CFR 1.27.

☐ Applicant certifies micro entity status. See 37 CFR 1.29.
Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

☐ A check in the amount of the fee is enclosed.

☒ Payment by credit card. ~~Form PTO-2038 is attached~~.

☐ The Director has already been authorized to charge fees in this application to a Deposit Account.

☒ The Director is hereby authorized to charge any fees which may be required, or credit any overpayment, to Deposit Account Number 23/2825 .

☒ Payment made via EFS-Web.

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

I am the

☐ applicant.

☐ attorney or agent of record. Registration number _____ .

☒ attorney or agent acting under 37 CFR 1.34. Registration number 52,728 .

/Tani Chen/	April 17, 2020
Signature	Date
Tani Chen, Sc.D.	617.646.8000
Typed or printed name	Telephone Number

NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications. Submit multiple forms if more than one signature is required, see below*.

☐ * Total of 1 forms are submitted.

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: April 17, 2020 Electronic Signature for Sarah Carletti: /Sarah Carletti/

FEE TRANSMITTAL

Complete if known	
Application Number	15/329,683-Conf. #7428
Filing Date	January 27, 2017
First Named Inventor	Xiaowei Zhuang
Examiner Name	F. W. Lu
Art Unit	1634
Practitioner Docket No.	H0498.70509US02

[X] Applicant asserts small entity status. See 37 CFR 1.27.

[ ] Applicant certifies micro entity status. See 37 CFR 1.29. Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

TOTAL AMOUNT OF PAYMENT ($)1,350.00

METHOD OF PAYMENT (check all that apply)

[ ] Check [X] Credit Card [ ] Money Order [ ] None [ ] Other (please identify): _____

[X] Deposit Account Deposit Account Number: 23/2825 Deposit Account Name: Wolf, Greenfield & Sacks, P.C.

For the above-identified deposit account, the Director is hereby authorized to (check all that apply):

[X] Charge fee(s) indicated below

[ ] Charge fee(s) indicated below, **except for the filing fee**

[X] Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17, **except for any excess claims fees or multiple dependent claim fee**

[ ] Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17

[X] Credit any overpayment of fee(s)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES (U = undiscounted fee; S = small entity fee; M = micro entity fee)

	FILING FEES			SEARCH FEES			EXAMINATION FEES			
Application Type	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**Fees Paid ($)**
Utility	300	150*	75	660	330	165	760	380	190	
Design	200	100	50	160	80	40	600	300	150	
Plant	200	100	50	420	210	105	620	310	155	
Reissue	300	150	75	660	330	165	2,200	1,100	550	
Provisional	280	140	70	0	0	0	0	0	0	

* The $150 small entity status filing fee for a utility application is further reduced to $75 for a small entity status applicant who files the application via EFS-Web.

2. EXCESS CLAIM FEES

Fee Description	**Undiscounted Fee ($)**	**Small Entity Fee ($)**	**Micro Entity Fee ($)**
Each claim over 20 (including Reissues)	100	50	25
Each independent claim over 3 (including Reissues)	460	230	115
Multiple dependent claims	820	410	205

Total Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**
_____ - 20 or HP =	_____ x	_____ =	_____

HP = highest number of total claims paid for, if greater than 20.

Multiple Dependent Claims	
Fee ($)	**Fee Paid ($)**
_____	_____

Indep. Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**
_____ - 3 or HP =	_____ x	_____ =	_____

HP = highest number of independent claims paid for, if greater than 3.

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper (excluding electronically filed sequence or computer listings under 37 CFR 1.52(e)), the application size fee due is $400 ($200 for small entity) ($100 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets	**Extra Sheets**	**Number of each additional 50 or fraction thereof**	**Fee ($)**	**Fee Paid ($)**
_____ - 100 =	_____ /50 =	_____ (round **up** to a whole number) x	_____ =	_____

4. OTHER FEE(S)

	Fees Paid ($)
Non-English specification, $130 fee (no small or micro entity discount)	
Non-electronic filing fee under 37 CFR 1.16(t) for a utility application, $400 fee ($200 small or micro entity)	
Other (e.g., late filing surcharge): 2253 Extension for response within third month	700.00
2801 Request for continued examination (RCE) - 1st ...	650.00

SUBMITTED BY

Signature	/Tani Chen/	Registration No. (Attorney/Agent)	52,728	Telephone	617.646.8000
Name (Print/Type)	Tani Chen, Sc.D.			Date	April 17, 2020

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: April 17, 2020 Electronic Signature for Sarah Carletti: /Sarah Carletti/

8077147.1

Electronic Patent Application Fee Transmittal

Application Number:	15329683
Filing Date:	27-Jan-2017
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Filer:	Tani Chen/Sarah carletti
Attorney Docket Number:	H0498.70509US02

Filed as Small Entity

Filing Fees for U.S. National Stage under 35 USC 371

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension - 3 months with $0 paid	2253	1	700	700
Miscellaneous:				
RCE- 1ST REQUEST	2801	1	650	650
		Total in USD ($)		**1350**

Electronic Acknowledgement Receipt

EFS ID:	39193265
Application Number:	15329683
International Application Number:	
Confirmation Number:	7428
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Customer Number:	86110
Filer:	Tani Chen/Sarah carletti
Filer Authorized By:	Tani Chen
Attorney Docket Number:	H0498.70509US02
Receipt Date:	17-APR-2020
Filing Date:	27-JAN-2017
Time Stamp:	15:37:51
Application Type:	U.S. National Stage under 35 USC 371

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$1350
RAM confirmation Number	E20204GF41126026
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Request for Continued Examination (RCE)	H049870509US02-RCE-TC.pdf	28147 d102c2f20eaced07cdfe5ff8a7d2b70e12ad83a6	no	1
Warnings:					
This is not a USPTO supplied RCE SB30 form.					
Information:					
2	Amendment Submitted/Entered with Filing of CPA/RCE	H049870509US02-116AMN-TC.pdf	49231 3c9d13f09c0fe4f63cfd87c6733c784b17dbec26	no	10
Warnings:					
Information:					
3	Extension of Time	H049870509US02-EXT-TC.pdf	24590 b73803f7c19a9bf3068e046bc0c25892d2c2154c	no	1
Warnings:					
Information:					
4	Fee Worksheet (SB06)	H049870509US02-FEE-TC.pdf	32852 6c46323b9c01cf77ff5cbc8ee6bee2327823d864	no	1
Warnings:					
Information:					
5	Fee Worksheet (SB06)	fee-info.pdf	32549 97cfec53c6ffa63d1bfa11d5f9db4177fcc980fb	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	167369		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Docket No.: H0498.70509US02
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	Xiaowei Zhuang
Application No.:	15/329,683
Confirmation No.:	7428
Filed:	January 27, 2017
For:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
Examiner:	F. W. Lu
Art Unit:	1634

Certificate of Electronic Filing under 37 C.F.R. § 1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: April 17, 2020 Electronic Signature for Sarah Carletti: /Sarah Carletti/

RESPONSE

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

INTRODUCTORY COMMENTS

In response to the Final Office Action dated October 18, 2019, reconsideration is respectfully requested in view of the following:

The Claims are reflected in the listing of claims which begins on page 2 of this paper.

Remarks begin on page 6 of this paper.

IN THE CLAIMS

Please replace all prior versions, and listings, of claims in the application with the following list of claims. Additions are indicated by underlining and deletions are indicate by strikeouts and/or double bracketing.

1. (Currently amended) A method, comprising:

(a) associating a plurality of nucleic acid targets ~~in a sample~~ with ~~a plurality of~~ valid codewords, comprising:

(i) defining a codeword space comprising a plurality of valid codewords, wherein the valid codewords in the codeword space ~~comprise a number~~ are defined by a series of positions and values ~~for each~~ at the positions, and the valid codewords form an error-checking and/or error-correcting code space; and

(ii) associating the valid codewords with a plurality of nucleic acids targets such that each nucleic acid target is associated with a unique valid codeword;

(b) exposing a sample suspected of containing at least some of the plurality of nucleic acid targets to a plurality of encoding nucleic acid probes that contain a representation of the valid codewords, comprising:

(i) providing a plurality of nucleic acid read sequences and associating a unique read sequence with each of the values, or a subset thereof, at each of the positions of the valid codewords;

(ii) ~~wherein at least some of the~~ creating a plurality of encoding nucleic acid probes, wherein each of the encoding nucleic acid probes comprises a first portion comprising a target sequence able to bind to one of the nucleic acid targets via nucleic acid hybridization, and a second portion comprising one or more read sequences taken from the plurality of read sequences~~, wherein each read sequence represents a value of a position within the codewords~~; and

(iii) exposing the sample to the plurality of encoding nucleic acid probes to cause the encoding nucleic acid probes to bind to at least some of the nucleic acid targets in the sample via nucleic acid hybridization; and

(c) ~~for each nucleic acid probe of the~~ identifying nucleic acid targets in the sample by determining valid codewords that correspond to the plurality of nucleic acid probes, ~~determining binding of the nucleic acid probe within the sample;~~ comprising:

~~creating codewords corresponding to the binding of the plurality of nucleic acid probes within the sample, wherein the values of the digits of the codewords are based on the read sequences present on the nucleic acid probes;~~

(i) exposing the sample to a first round of one or more readout nucleic acid probes, wherein at least some of the readout nucleic acid probes comprise a first portion and a second portion, the first portion comprising a sequence substantially complementary to one of the read sequences, and the second portion comprising a fluorescent signaling entity;

(ii) imaging the sample to determine a fluorescence signal of the fluorescent signaling entity in at least some locations in the sample;

(iii) determining read sequences and codewords at the locations based on the fluorescence signal;

(iv) ~~for at least some of the codewords,~~ matching the codewords to [[a]] the valid codewords in the codeword space wherein, if no match is found, either discarding the codeword or applying error correction to the codeword to ~~form a~~ match one of the valid codewords~~, the valid codewords being a plurality of codewords assigned to the plurality of the nucleic acid targets~~; and

~~determining a nucleic acid abundance and/or spatial distribution within the sample using the valid codewords corresponding to the binding of the plurality of nucleic acid probes within the sample~~

(v) identifying nucleic acid targets in the sample based on the codewords matched to the valid codewords.

2-33. (Cancelled)

34. (Currently amended) The method of claim ~~31~~ 1, further comprising inactivating at least some of the ~~first~~ fluorescent signaling entity and/or removing at least some of the ~~first~~

fluorescent signaling entity from the sample ~~prior to exposing the sample to the second secondary probe~~ after exposing the sample to one or more readout nucleic acid probes and imaging the sample to determine the fluorescence signal of the fluorescent signaling entities.

35-70. (Cancelled)

71. (Currently amended) The method of claim 1, comprising determining ~~binding of the nucleic acid probes within the sample~~ the fluorescent signaling entity at a resolution better than 300 nm.

72-74. (Cancelled)

75. (Previously presented) The method of claim 1, wherein the sample comprises a cell.

76. (Cancelled)

77. (Previously presented) The method of claim 75, wherein the cell is fixed.

78. (Currently amended) The method of claim 1, comprising determining ~~binding of the nucleic acid probes~~ the fluorescent signal entity by imaging at least a portion of the sample using a light microscope.

79-176. (Cancelled)

177. (New) The method of claim 1, further comprising:

exposing the sample to a second round of one or more readout nucleic acid probes, wherein at least some of the readout nucleic acid probes of the second round comprise a first portion and a second portion, the first portion comprising a sequence substantially complementary to one of the read sequences different from the read sequences of the first

round, and the second portion comprising a fluorescent signaling entity; and

imaging the sample to determine a fluorescence signal of the fluorescent signaling entity of the second round of one or more readout nucleic acid probes in at least some locations in the sample.

178. (New) The method of claim 1, wherein the codeword position values are binary values.

179. (New) The method of claim 1, further comprising determining an abundance and/or spatial distribution of the nucleic acid targets.

180. (New) The method of claim 1, further comprising removing unbound readout nucleic acid probes prior to imaging the sample.

181. (New) The method of claim 34, wherein the acts of exposing the sample to one or more readout nucleic acid probes, imaging the sample, and inactivating at least some of the fluorescent signaling entities on the read out probes are performed for each of the values, or subset thereof, at each of the positions within the codeword space.

182. (New) The method of claim 34, wherein the acts of exposing the sample to one or more readout nucleic acid probes, imaging the sample, inactivating at least some of the fluorescent signaling entities on the readout probes, and assigning codeword position values are performed for each of the values, or subset thereof, at each of the positions within the codeword space.

REMARKS

The undersigned thanks the Examiner for the courtesy of a telephone conducted on January 13, 2020 with the undersigned and Prof. Jeffrey Moffitt, one of the inventors of the present application. The present response addresses substantive points discussed during the interview. For example, claim 1 has been amended to better clarify the codewords, steps taken to produce codewords, the meaning of valid codewords, how the valid codewords are associated with the nucleic acid targets, how the sample suspected of containing nucleic acid targets are exposed to encoding nucleic acid probes that can both bind to the nucleic acid targets and contain a representation of the valid codewords, and how the nucleic acid targets in the sample are identified by determining the valid codewords that correspond to the encoding probes bound to nucleic acids. Many of the dependent claims have been correspondingly amended. In addition, claims 7, 10, 18, 26, 31, 44, 50, 95, 96, 99, 103, 122, 126, and 129 have been cancelled without prejudice. Furthermore, new dependent claims 177-182 have been added supported by the specification. Accordingly, the present response is believed to constitute a complete written statement of the reasons presented at the interview as warranting favorable action, as required by 37 C.F.R. § 1.133(b).

No new matter has been added. Claims 1, 34, 71, 75, 77, 78, and 177-182 are now pending for examination.

Claim 1

For the sake of clarity, a clean form of claim 1, without strikeouts and underlining, is presented here:

1. A method, comprising:

(a) associating a plurality of nucleic acid targets with valid codewords, comprising:

(i) defining a codeword space comprising a plurality of valid codewords, wherein the valid codewords in the codeword space are defined by a series of positions and values at the positions, and the valid codewords form an error-checking and/or error-correcting code space; and

(ii) associating the valid codewords with a plurality of nucleic acids targets such that each nucleic acid target is associated with a unique valid codeword;

(b) exposing a sample suspected of containing at least some of the plurality of nucleic acid targets to a plurality of encoding nucleic acid probes that contain a representation of the valid codewords, comprising:

(i) providing a plurality of nucleic acid read sequences and associating a unique read sequence with each of the values, or a subset thereof, at each of the positions of the valid codewords;

(ii) creating a plurality of encoding nucleic acid probes, wherein each of the encoding nucleic acid probes comprises a first portion comprising a target sequence able to bind to one of the nucleic acid targets via nucleic acid hybridization, and a second portion comprising one or more read sequences taken from the plurality of read sequences; and

(iii) exposing the sample to the plurality of encoding nucleic acid probes to cause the encoding nucleic acid probes to bind to at least some of the nucleic acid targets in the sample via nucleic acid hybridization; and

(c) identifying nucleic acid targets in the sample by determining valid codewords that correspond to the plurality of nucleic acid probes, comprising:

(i) exposing the sample to a first round of one or more readout nucleic acid probes, wherein at least some of the readout nucleic acid probes comprise a first portion and a second portion, the first portion comprising a sequence substantially complementary to one of the read sequences, and the second portion comprising a fluorescent signaling entity;

(ii) imaging the sample to determine a fluorescence signal of the fluorescent signaling entity in at least some locations in the sample;

(iii) determining read sequences and codewords at the locations based on the fluorescence signal;

(iv) matching the codewords to the valid codewords in the codeword space wherein, if no match is found, either discarding the codeword or applying error correction to the codeword to match one of the valid codewords; and

(v) identifying nucleic acid targets in the sample based on the codewords matched to the valid codewords.

Claim Objections

Claims 1, 31, and 50 has been objected to for various formalities.

Based on the above-described amendments to claim 1, it is believed that the objections to claim 1 are now moot. Also, claims 31 and 50 have each been cancelled, rendering their objections moot. Accordingly, for at least these reasons, withdrawal of these objections is respectfully requested.

Rejections under 35 U.S.C. § 101

Claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 have been rejected under 35 U.S.C. § 101 because the claimed invention is directed to a judicial exception (i.e., a law of nature, a natural phenomenon, or an abstract idea) without significantly more.

During the January 13, 2020 interview, several clarifying amendments were suggested by the Examiner in order to address the § 101 rejections, including how the codewords were experimentally constructed. In response, independent claim 1 has been amended and clarified to recite associating a plurality of nucleic acid targets with valid codewords, exposing a sample suspected of containing at least some of the plurality of nucleic acid targets to a plurality of encoding nucleic acid probes that contain a representation of the valid codewords, and identifying nucleic acid targets in the sample by determining the valid codewords that correspond to the plurality of encoding nucleic acid probes, as well as detailed explanation of how each of these steps are performed, as shown above.

Thus, for at least these reasons, withdrawal of these rejections is respectfully requested. In addition, it is not conceded that the rejection of the claims as pending prior to amendment herein under § 101 were proper.

Rejections under 35 U.S.C. § 112, ¶ 1 and ¶ 2

Claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 have been rejected under 35 U.S.C. § 112, ¶ 1, because the specification, while being enabling for exposing a sample to a plurality of nucleic acid probes, does not reasonably provide enablement for performing the methods recited in claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78. In addition, claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 have been rejected 35 U.S.C. § 112, ¶ 2, for being indefinite for failing to particularly point out and distinctly claim the subject matter which the applicant regards as the invention.

During the January 13, 2020 interview, the Examiner suggested that amendments to better clarify the codewords and the steps taken to produce codewords experimentally would address the 35 U.S.C. § 112, ¶ 1 rejections concerning enablement and the ¶ 2 rejections concerning definiteness. In response, independent claim 1 has been amended as noted above. Thus, for at least these reasons, withdrawal of these rejections is respectfully requested. In addition, it is not conceded that the rejections of the claims as pending prior to amendment herein under § 112, ¶ 1 and ¶ 2 were proper.

Rejections under 35 U.S.C. § 112, ¶ 4

Claim 10 has been rejected 35 U.S.C. § 112, ¶ 4, as being of improper dependent form for failing to further limit the subject matter of the claim upon which it depends, or for failing to include all of the limitations of the claim upon which it depends.

Claim 10 has been cancelled, thereby rendering this rejection moot. Withdrawal of this rejection is thus respectfully requested. In addition, it is not conceded that this rejection was proper.

CONCLUSION

In view of the foregoing remarks, this application should now be in condition for allowance. A notice to this effect is respectfully requested. If the Examiner believes, after the foregoing remarks, that the application is not in condition for allowance, the Examiner is requested to call the undersigned at 617-646-8000.

If this response is not considered timely filed and if a suitable request for an extension of time is otherwise absent, any necessary extension of time is hereby requested. If there is a fee occasioned by this response, including an extension fee, the Director is hereby authorized to charge any deficiency or credit any overpayment in the fees filed, asserted to be filed or which should have been filed herewith to Deposit Account No. 23/2825, under Docket No. H0498.70509US02 from which the undersigned is authorized to draw.

Dated: April 17, 2020

Respectfully submitted,

Electronic signature: /Tani Chen/
Tani Chen, Sc.D.
Registration No.: 52,728
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

FORM PTO-1449/A and B (modified PTO/SB/08)				APPLICATION NO.: 15/329,683	ATTY. DOCKET NO.: H0498.70509US02
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				FILING DATE: January 27, 2017	CONFIRMATION NO.: 7428
				FIRST NAMED INVENTOR: Xiaowei Zhuang	
Sheet	1	of	1	GROUP ART UNIT: 1634	EXAMINER: F. W. Lu

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document Number	Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		6,150,516	A1	Brenner et al.	11-21-2000
		7,226,734	B2	Chee et al.	06-05-2007
		2006-0275782	A1	Gunderson et al.	12-07-2006
		2015-0267251	A1	Cai et al.	09-24-2015
		2019-0233812	A1	Zhuang et al.	08-01-2019
		2019-0264270	A1	Zhuang et al.	08-29-2019
		2019-0276881	A1	Zhuang et al.	09-12-2019
		2020-0095630	A1	Zhuang et al.	03-26-2020

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document Office/ Country	Number	Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)
		WO	2018/026873	A1	California Institute of Technology	02-08-2018	

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		Chinese Office Action mailed March 5, 2020 for Application No. CN 201580051029.5 (H0498.70510CN00).	
		International Preliminary Report on Patentability mailed December 5, 2019 for Application No. PCT/US18/34651 (H0498.70637WO00).	
		Invitation to Pay Additional Fees for Application No. PCT/US2019/065857 mailed March 3, 2020 (H0498.70680WO00).	
		EPSTEIN et al., Reutilization of Previously Hybridized Slides for Fluorescence In Situ Hybridization. Cytometry. 1995;21:378-381.	
		GUNDERSON et al., Decoding Randomly Ordered DNA Arrays. Methods. 2004;14:870–877.	

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

EXAMINER:	DATE CONSIDERED:

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

8134927.1

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
08 February 2018 (08.02.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/026873 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01) *G01N 21/64* (2006.01)

(21) International Application Number:
PCT/US2017/044994

(22) International Filing Date:
01 August 2017 (01.08.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

15/225,820	01 August 2016 (01.08.2016)	US
15/298,219	19 October 2016 (19.10.2016)	US
62/428,910	01 December 2016 (01.12.2016)	US
62/456,291	08 February 2017 (08.02.2017)	US
62/523,127	21 June 2017 (21.06.2017)	US

(71) Applicant: **CALIFORNIA INSTITUTE OF TECHNOLOGY** [US/US]; 1200 E. California Blvd., MC 114-96, Pasadena, CA 91125 (US).

(72) Inventors: **CAI, Long**; c/o California Institute Of Technology, 1200 E. California Blvd., MC 127-72, Pasadena, CA 91125 (US). **ENG, Chee, Huat**; c/o California Institute Of Technology, 1200 E. California Blvd., MC 127-72, Pasadena, CA 91125 (US). **SHAH, Sheel**; c/o California Institute Of Technology, 1200 E. California Blvd., MC 127-72, Pasadena, CA 91125 (US).

(74) Agent: **WANG, Stephanie**; DLA Piper LLP US, 2000 University Avenue, East Palo Alto, CA 94303 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

(54) Title: SEQUENTIAL PROBING OF MOLECULAR TARGETS BASED ON PSEUDO-COLOR BARCODES WITH EMBEDDED ERROR CORRECTION MECHANISM

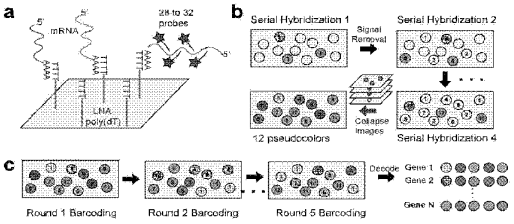


FIG. 4

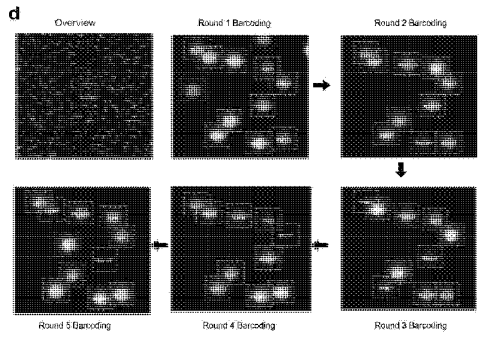


FIG. 4 (continued)

(57) Abstract: The present invention, among other things, provides technologies for detecting and/or quantifying nucleic acids in cells, tissues, organs or organisms. Pre-designed barcodes are associated specific molecular targets through sequential hybridization experiments. A pseudo¬color based barcoding scheme is developed to overcome limitations in the previous generation of the technology such as lack of visual signals that can be associated with the probes or small internal within cell when carrying out in situ experiments.. The current method can be applied to both in vitro and in situ analysis. According to the method, each barcoding round comprises multiple serial hybridizations where a small number of colored signals (that are associated with probes) are used in each hybridization experiment within a serial hybridization round. Images from each serial hybridization experiment within the same serial hybridization round are combined to form a composite image for each barcoding round. In each barcoding round, the same set of molecular targets are analyzed. After all barcoding rounds are completed, associated of the barcode with these molecular targets is completed.

[Continued on next page]

WO 2018/026873 A1

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

SEQUENTIAL PROBING OF MOLECULAR TARGETS BASED ON PSEUDO-COLOR BARCODES WITH EMBEDDED ERROR CORRECTION MECHANISM

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to United States Patent Application No. 15/225,820, filed on August 1, 2016 and entitled "Multiplex Labeling of Molecules by Sequential Hybridization Barcoding Using Probes With Cleavable Linkers," United States Patent Application No. 15/298,219, filed on October 19, 2016 and entitled "Error Correction of Multiplex Imaging Analysis by Sequential Hybridization," United States Patent Provisional Application No. 62/428,910, filed on December 1, 2016 and entitled "Single Molecule Profiling Through Serial and Barcoded Hybridization," United States Patent Provisional Application No. 62/456,291, filed on February 8, 2017 and entitled "Imaging-based Transcriptomic and Translational Profiling of 1000 Genes with *in vitro* seqFISH," and United States Patent Provisional Application No. 62/523,127, filed on June 21, 2017 and entitled "Transcriptome Profiling of 10,000 mRNAs by RNA SPOTs," each of which is hereby incorporated by reference herein in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0002] This invention was made with government support under Grant No. HD075605 and under Grant No. OD008530 awarded by the National Institutes of Health. The government has certain rights in the invention.

BACKGROUND OF THE INVENTION

[0003] Transcription profiling of cells are essential for many purposes. Microscopy imaging which can resolve multiple mRNAs in single cells can provide valuable information regarding transcript abundance and localization, which are important for understanding the molecular basis of cell identify and developing treatment for diseases. Molecular profiling such as transcriptomic profiling of biological samples is essential for various purposes. For example, it would allow one to assess gene expression levels to detect and identify abnormal growth states such as cancers. Using nucleic acid detection as an example, current nucleic acid-based assays

such as qPCR and microarrays have been useful, but they do not reach single molecule sensitivity. Next generation sequencing, on the other hand, involves amplification of the sample and reverse transcription of mRNA which can introduce biases and inaccuracies in quantification. Moreover, sample preparation and sequencing can be time-consuming and economically costly. Despite the fact that imaging has been used for mRNA transcripts quantification, it is limited to a few hundreds of genes. Many scientific questions require thousands of genes or even the whole transcriptome to be quantified.

[0004] What is needed in the are better methods and systems for carrying out imaging based transcriptomic profiling at a single molecule sensitivity with high accuracy in a time efficient manner.

SUMMARY OF THE INVENTION

[0005] The present invention provides certain insights into challenges or defects associated with existing technologies for profiling transcripts or DNA loci in cells, particularly for single cells. Moreover, the present invention provides new technologies for achieving effective such profiling, including of single cells. Provided technologies are broadly useful, including for example for profiling of isolated cells, cells in tissues, cells in organs, and/or cells in organisms.

[0006] For example, the present invention provides the insight that existing technologies such as single cell RNA-seq or qPCR require single cells to be isolated and put into multi-well format, which is a multiple step process that can be cost prohibitive, labor intensive and prone to artifacts. Furthermore, the present invention recognizes that existing *in situ* sequencing technologies that use enzymatic reactions to convert the mRNA into a DNA template first can be highly inefficient (for example in the mRNA to DNA conversion process), so that, often, only a small fraction of the RNAs are converted and detected. The present invention provides the particular insight that one major downside of such low efficiency, which is estimated at 1% for RT and 10% for PLA, is that it can introduce significant noise ad bias in the gene expression measurements. The present invention further recognizes that existing spectral mRNA barcoding technologies that utilize single molecule fluorescence in situ hybridization (smFISH) require distinct fluorophores for scale up, and may be limited in the number of barcodes that can be

generated. smFISH also requires splitting probes into barcoding subsets during hybridization. Because smFISH often uses two or more colors for a target, it produces high density of objects in the image, which can increase the complexity of data analysis.

[0007] Among other things, the present inventions provides new technologies for profiling, for example, transcripts and/or DNA loci, that overcome one or more or all of the problems associated with methods prior to the present invention. In some embodiments, the present invention provides methods for detecting multiple targets, *e.g.*, transcripts or DNA loci, in a cell through a sequential barcoding scheme that permits multiplexing of different targets.

[0008] In one aspect, disclosed herein is a method of barcoding molecular targets. The method comprises: identifying N molecular targets in a biological sample, wherein the N molecular targets are immobilized; associating a unique barcode to each molecular target via n sequential barcoding rounds (where $n \geq 2$), wherein each barcoding round comprises m serial hybridizations of probes collectively bound to the N molecular targets (where $m \geq 2$), and optionally removing probes between two barcoding rounds. In some embodiments, each serial hybridization in turn further comprises: contacting one or more groups of probes to a subset of the N molecular targets, the total number of groups of probes corresponding to the number of molecular targets in the subset, where probes in each group comprise one or more binding sequences specifically targeting a molecular target in the subset, where each probe is capable of generating at least one detectable visual signal representing binding of the probe to a molecular target in the subset, and where probes in the one or more groups generate one or more different detectable visual signals corresponding to the number of molecular targets in the subset; detecting the detectable visual signals that reflect the binding between the one or more groups of probes and the subset of the N molecular targets; and removing the visual signals, when applicable, prior to the next serial hybridization; wherein the unique barcode to each molecular target consists of n components, each component is assigned from S unique symbols, where S is an integer that equal to or greater than $\sqrt[n]{N}$. N and n are both integer.

[0009] In some embodiments, the detecting the detectable visual signals comprises: capturing, for each serial hybridization round, an image of the detectable visual signals that

reflect the binding between the one or more groups of probes and the subset of the N molecular targets.

[0010] In some embodiments, the method further comprises: generating, for each barcoding round, a composite image by superimposing m images corresponding to the m serial hybridizations, wherein the m images are aligned based on one or more alignment references whose positions remain constant relative to the biological sample.

[0011] In some embodiments, the method further comprises: applying Gaussian analysis to super-localize the detectable visual signals in an image.

[0012] In some embodiments, the method further comprises: decoding the detectable visual signals in each composite image based on the unique barcodes for the N molecular targets and the S unique symbols.

[0013] In some embodiments, the method further comprises: detecting reference visual signals associated with the one or more alignment reference.

[0014] In some embodiments, the one or more alignment references comprise one or more selected from the group consisting of an oligonucleotide sequence immobilized on the coverslips and detected by a complementary oligo, a common sequence embedded in all probes, a microscopic object, a metal bead, a gold bead, a polystyrene bead, a PCR handle sequence on a primary binding probe, and combinations thereof.

[0015] In some embodiments, the n sequential barcoding rounds includes x round for error correction, where x is an integer equal or greater than 1; and wherein assigning unique barcodes for each of N molecular targets requires S unique symbols, where S is an integer equal or greater than $\sqrt[n-x]{N}$.

[0016] In some embodiments, the biological sample comprises a tissue sample, a cell sample, a cell extract sample, a nucleic acid sample, an RNA transcript sample, a protein sample, an mRNA sample, DNA molecules, protein molecules, RNA and DNA isoform molecules, single nucleotide polymorphism molecules, or combinations thereof.

[0017] In some embodiments, the method further comprises: determining a secondary molecular target that are associated with the N molecular targets by contacting the biological sample with molecules specifically binding to the secondary molecular target.

[0018] In some embodiments, the secondary molecular target comprises one selected from the group consisting of a RNA binding protein molecule, ribosome, a DNA binding protein molecule, a transcription factor, a chromatin binding protein, a protein binding molecule, a scaffold protein, and combinations thereof.

[0019] In some embodiments, probes in the one or more groups of probes further comprise: one or more binding sequences each specifically targeting one or more sites within a molecular target in the biological sample; and n unique readout sequences associated with the one or more binding sequences, wherein, in each barcoding round, only one unique readout sequence is associated with a detectable visual signal for a particular molecular target.

[0020] In some embodiments, the one or more binding sequences target multiple different sites within the same molecular target. In some embodiments, the one or more binding sequences target multiple different sites within different molecular targets.

[0021] In some embodiments, each probe comprises one or more of the *n* unique readout sequences.

[0022] In some embodiments, at least one of the *n* unique readout sequences is located in an overhang sequence directly connected to the binding sequence of a probe.

[0023] In some embodiments, at least one of the n unique readout sequences is indirectly connected to the binding sequence of a probe via one or more intermediate molecules.

[0024] In some embodiments, the one or more intermediate molecules comprise an RNA bridge probe, a DNA bridge probe, a protein bridge probe, a probe for hybridization chain reaction (HCR), a hairpin nucleic acid probe, an HCR initiator, an HCR polymer, or combinations thereof.

[0025] In some embodiments, the one or more binding sequences specifically target one or more non-nucleic acid sites in the molecular target, and wherein the n unique readout sequences comprising nucleic acid sequences that are directly or indirectly connected to the binding sequences of the probes.

[0026] In some embodiments, the detectable visual signal is connected to the binding sequence of a probe or an intermediate molecule via a cleavable linker.

[0027] In some embodiments, the one or more binding sequences comprises a peptide sequence binding to a specific antigen within a particular molecular target, an aptmer, or click chemistry group.

[0028] In some embodiments, the S unique symbols comprise colors, numbers, letters, shapes, or combinations thereof.

[0029] In some embodiments, for each serial hybridization, the one or more groups of probes to a non-overlapping subset of the *N* molecular targets.

[0030] In one aspect, disclosed herein is a method of hybridization analysis of binding between labeled probes and molecular targets in a biological sample. The method comprises: generating multiple composite images of labeled probes bound to a plurality of molecular targets in the biological sample, wherein each composite image is generated from a plurality of images of labeled probes collectively bound to the plurality of molecular targets, wherein the plurality of molecular targets are immobilized within the biological sample, and wherein each image of the plurality of images reveals: labeled probes bound to a subset of molecular targets within the plurality of molecular targets, wherein the labeled probes comprise one or more groups of probes, the total number of groups of probes corresponding to the number of molecular targets in the subset, wherein probes in each group comprise one or more binding sequences specifically targeting a molecular target in the subset, and wherein each labeled probe is capable of generating a visual signal representing binding of the probe to a molecular target; and one or more reference targets whose positions remain constant in the biological sample for aligning the plurality of images.

[0031] In some embodiments, the biological sample comprises a tissue sample, a cell sample, a cell extract sample, a nucleic acid sample, an RNA transcript sample, a protein sample, an mRNA sample, DNA molecules, protein molecules, RNA and DNA isoform molecules, single nucleotide polymorphism molecules, or combinations thereof.

[0032] In some embodiments, in each image, the labelled probes bind to a non-overlapping subset of molecular targets within the plurality of molecular targets.

[0033] In some embodiments, the method further comprises: contacting the one or more groups of probes with the subset of molecular targets of the plurality of molecular targets; detecting visual signals that reflect the binding between the one or more groups of probes and

molecular targets in the subset; and removing the visual signals, when applicable, prior to a next round of hybridization of labeled probes binding to a new subset of molecular targets within the plurality of molecular targets.

[0034] In some embodiments, the method further comprises: detecting reference visual signals associated with the one or more alignment references.

[0035] In some embodiments, the method further comprises: aligning the plurality of images based on the positions of the one or more alignment references.

[0036] In one aspect, disclosed herein is a sequential hybridization method that comprises the steps of: identifying a plurality of target genes; and associating, via sequential hybridization of binding probes to the plurality of target genes, a first plurality of unique codes with the plurality of target genes, where each target gene in the plurality of target genes is represented by a unique code in the first plurality of unique codes, where the sequential hybridization comprises n rounds of hybridization (where $n \geq 2$). Here, each round of hybridization in n rounds of hybridization in turn comprises the steps of contacting the plurality of target genes with a plurality of binding probes, where each probe in the plurality of binding probes comprises: a binding sequence that specifically binds a target sequence in a gene in the plurality of target genes, where target genes from the plurality of target genes are spatially transfixed from each other, and where each probe is capable of emitting a detectable visual signal upon binding of the probe to a target sequence; detecting visual signals that reflect the binding between the plurality of binding probes and the plurality of target genes; and removing the visual signals, when applicable, prior to the next round of hybridization. In some embodiments, probes used in the n rounds of hybridization are capable of emitting at least F types of detectable visual signals (where $F \geq 2$ and F^n is greater than the number of target genes in the plurality of target genes). In some embodiments, a unique code in the first plurality of unique codes for a target gene consists of n components. In some embodiments, each component is determined by visual signals that reflect the binding between binding probes and the target gene during one of the n rounds of hybridization. In some embodiments, the n rounds of hybridization include m error correction round ($m \geq 1$). In some embodiments, a second plurality of unique codes for the plurality of target genes is generated after the m error correction round is removed from the n

rounds of hybridization. In some embodiments, each unique code in the second plurality of unique codes consists of ($n - m$) components and uniquely represents a target gene in the plurality of target genes.

[0037] In some embodiments, the plurality of target genes are located on immobilized nucleic acids selected from the group consisting of mRNAs, chromosomal DNAs and combinations thereof. In some embodiments, n is 4 or greater, 5 or greater, or 10 or greater. In some embodiments, the m error correction round comprises one round of the n rounds of hybridization. In some embodiments, the one round of the n rounds of hybridization is a repeat of one of the remaining one or more ($n - 1$) rounds of the n rounds of hybridization. In some embodiments, where $m \leq 0.5n$.

[0038] In some embodiments, the at least F types of detectable visual signals comprises one selected from the group consisting of a fluorescence signal, a color signal, a red signal, a green signal, a yellow signal, a combined color signal representing two or more colors, and combinations thereof.

[0039] In some embodiments, a probe in the plurality of binding probes further comprises a signal moiety that emits a detectable visual signal upon binding of the probe to a target sequence.

[0040] In some embodiments, the signal moiety is connected to the binding sequence of the probe via a cleavable linker.

[0041] In some embodiments, each component of a n-component unique code in the first plurality of unique codes is assigned a numerical value that corresponds to one of the at least F types of detectable visual signals; and where at least one component of the n-component unique code is determined based on the numerical values of all or some of the other n-1 components. In some embodiments, the n-component unique code is determined as:

$\{j_1, j_2, \ldots (a_1*j_1 + a_2*j_2 \ldots + a_n*j_n + C) \text{mod } F, \ldots, j_n\}$, where j_1 is a numerical value that corresponds the detectable visual signals used in the first round of hybridization, j_2 is a numerical value that corresponds the detectable visual signals used in the second round of hybridization, and j_n is a numerical value that corresponds the detectable visual signals used in the nth round of hybridization; and where $j_1, j_2, \ldots j_n$, $a_1, a_2, \ldots a_n$ and n are none zero integers and C is an integer.

[0042] In some embodiments, m, n, F, i, j and k are all integers.

[0043] In one aspect disclosed herein is a hybridization method that comprises the steps of: identifying a plurality of target genes; performing sequential hybridization of binding probes to the plurality of target genes, where the sequential hybridization comprises n rounds of hybridization (where $n \geq 2$). Here, each round of hybridization in n rounds of hybridization in turn comprises: contacting the plurality of target genes with a plurality of binding probes, where each probe in the plurality of binding probes comprises: a binding sequence that specifically binds a target sequence in a gene in the plurality of target genes, where target genes from the plurality of target genes are spatially transfixed from each other, and where each probe is capable of emitting a detectable visual signal upon binding of the probe to a target sequence; detecting visual signals that reflect the binding between the plurality of binding probes and the plurality of target genes, where each target gene in the plurality of target genes is represented by visual signals that are unique for the target gene, and where probes used in the n rounds of hybridization are capable of emitting at least F types of detectable visual signals (where $F \geq 2$, and F^n is greater than the number of target genes in the plurality of target genes); and removing the visual signals, when applicable, prior to the next round of hybridization; and performing serial hybridizations against one or more serial target genes, where the expression level of each serial target gene is above a predetermined threshold value, and where each serial hybridization in turn comprises: contacting the one or more serial target genes with a plurality of binding probes, where each probe in the plurality of binding probes comprises: a binding sequence that specifically binds a target sequence in a serial target gene in the one or more serial target genes, where one or more serial target genes are spatially transfixed from each other, where each probe is capable of emitting a detectable visual signal upon binding of the probe to the target sequence, and where probes binding to target sequences in the same serial target gene emit the same detectable visual signals; and detecting visual signals that reflect the binding between the plurality of binding probes and the one or more serial target gene.

[0044] In some embodiments, the n rounds of hybridization generate a first plurality of unique codes, where each target gene in the plurality of target genes is represented by a unique code in the first plurality of unique codes.

[0045] In some embodiments, where a unique code in the first plurality of unique codes for a target gene consists of n components, and where each component is determined by visual

signals that reflect the binding between binding probes and the target gene during one of the n rounds of hybridization.

[0046] In some embodiments, the *n* rounds of hybridization include *m* error correction round ($m \geq 1$), where a second plurality of unique codes for the plurality of target genes is generated after the *m* error correction round is removed from the *n* rounds of hybridization, and where each unique code in the second plurality of unique codes consists of $(n - m)$ components and uniquely represents a target gene in the plurality of target genes.

In some embodiments, the method of hybridization further comprises the step of: identifying the one or more serial target genes based on expression levels of candidate target genes.

In some embodiments, the plurality of target genes are located on immobilized nucleic acids selected from the group consisting of mRNAs, chromosomal DNAs and combinations thereof.

[0047] In some embodiments, the one or more serial target genes are located on immobilized nucleic acids selected from the group consisting of mRNAs, chromosomal DNAs and combinations thereof.

[0048] In some embodiments, each unique code in the first plurality of unique codes consists of n component, where each component of a *n*-component unique code in the first plurality of unique codes is assigned a numerical value that corresponds to one of the at least *F* types of detectable visual signals; and where at least one component of the *n*-component unique code is determined based on the numerical values of all or some of the other *n*-1 components.

In some embodiments, the *n*-component unique code is determined as:

$\{j_1, j_2, \ldots (a_1*j_1 + a_2*j_2 \ldots + a_n*j_n + C) \bmod F, \ldots, j_n\}$, where j_1 is a numerical value that corresponds the detectable visual signals used in the first round of hybridization, j_2 is a numerical value that corresponds the detectable visual signals used in the second round of hybridization, and j_n is a numerical value that corresponds the detectable visual signals used in the *n*th round of hybridization; and where $j_1, j_2, \ldots j_n, a_1, a_2, \ldots a_n$ are none zero integers and C is an integer.

[0049] In one aspect, disclosed herein is a non-transitory computer-readable medium containing instructions that, when executed by a computer processor, cause the computer processor to: associate, via sequential hybridization of binding probes to a plurality of target genes, a first plurality of unique codes with the plurality of target genes, where each target gene in the plurality of target genes is represented by a unique code in the first plurality of unique codes,

where the sequential hybridization comprises n rounds of hybridization (where $n \geq 2$). Here each round of hybridization in n rounds of hybridization in turn comprises: contacting the plurality of target genes with a plurality of binding probes, where each probe in the plurality of binding probes comprises: a binding sequence that specifically binds a target sequence in a gene in the plurality of target genes, where target genes from the plurality of target genes are spatially transfixed from each other, and where each probe is capable of emitting a detectable visual signal upon binding of the probe to a target sequence; detecting visual signals that reflect the binding between the plurality of binding probes and the plurality of target genes; and removing the visual signals, when applicable, prior to the next round of hybridization.

[0050] In some embodiments, probes used in the n rounds of hybridization are capable of emitting at least F types of detectable visual signals (where $F \geq 2$ and F^n is greater than the number of target genes in the plurality of target genes). In some embodiments, a unique code in the first plurality of unique codes for a target gene consists of n components. In some embodiments, each component is determined by visual signals that reflect the binding between binding probes and the target gene during one of the n rounds of hybridization. In some embodiments, the n rounds of hybridization include m error correction round ($m \geq 1$). In some embodiments, a second plurality of unique codes for the plurality of target genes is generated after the m error correction round is removed from the n rounds of hybridization. In some embodiments, each unique code in the second plurality of unique codes consists of $(n - m)$ components and uniquely represents a target gene in the plurality of target genes.

[0051] In one aspect, disclosed herein is a non-transitory computer-readable medium containing instructions that, when executed by a computer processor, cause the computer processor to: perform sequential hybridization of binding probes to a plurality of target genes, where the sequential hybridization comprises n rounds of hybridization (where $n \geq 2$).

[0052] Here, each round of hybridization in n rounds of hybridization comprises: contacting the plurality of target genes with a plurality of binding probes, where each probe in the plurality of binding probes comprises: a binding sequence that specifically binds a target sequence in a gene in the plurality of target genes, where target genes from the plurality of target genes are spatially transfixed from each other, and where each probe is capable of emitting a detectable visual signal upon binding of the probe to a target sequence; detecting visual signals

that reflect the binding between the plurality of binding probes and the plurality of target genes, where each target gene in the plurality of target genes is represented by visual signals that are unique for the target gene, and where probes used in the n rounds of hybridization are capable of emitting at least F types of detectable visual signals (where $F \geq 2$, and F^n is greater than the number of target genes in the plurality of target genes); and removing the visual signals, when applicable, prior to the next round of hybridization; and perform serial hybridizations against one or more serial target genes, where the expression level of each serial target gene is above a predetermined threshold value, where each serial hybridization comprises: contacting the one or more serial target genes with a plurality of binding probes, where each probe in the plurality of binding probes comprises: a binding sequence that specifically binds a target sequence in a serial target gene in the one or more serial target genes, where one or more serial target genes are spatially transfixed from each other, where each probe is capable of emitting a detectable visual signal upon binding of the probe to the target sequence, and where probes binding to target sequences in the same serial target gene emit the same detectable visual signals; and detecting visual signals that reflect the binding between the plurality of binding probes and the one or more serial target gene.

[0053] In any of the embodiments disclosed herein, m, n, F, i, j and k are all integers. Embodiments disclosed herein can be applied individually or in combination in any aspect disclosed herein.

[0054] In one aspect, disclosed herein is a composition comprising a plurality of primary probes, a first plurality of bridge probes, and first plurality of readout probes.

[0055] In some embodiments, each primary probe in the plurality of primary probes comprises: a primary binding sequence that binds to a complementary target sequence in a target nucleic acid molecule, and a first overhang sequence connected to one end of the primary binding sequence.

[0056] In some embodiments, each bridge probe in the first plurality of bridge probes comprises a binding sequence that specifically binds to all or a part of the first overhang sequence of a primary probe of the plurality of primary probes, and one or more readout binding targets connected in series and linked to the binding sequence.

[0057] In some embodiments, each readout probe in the first plurality of readout probes comprises: a readout binding sequence that specifically binds to a first readout binding target of the one or more readout binding targets of a bridge probe of the first plurality of bridge probes, and a signal moiety linked to the readout binding sequence via a cleavable linker.

[0058] In these embodiments, the signal moiety is capable of emitting a first detectable visual signal upon binding of each readout probe from the first plurality of readout probes to the first readout binding target of one of the one or more readout binding targets.

[0059] In some embodiments, the composition further comprises: a second plurality of readout probes, wherein each readout probe comprises: a readout binding sequence that specifically binds to a second readout binding target of the one or more readout binding targets in a bridge probe of the first plurality of bridge probes, and a signal moiety linked to the readout binding sequence via a cleavable linker.

[0060] In these embodiments, the signal moiety is capable of emitting a second detectable visual signal upon binding of each readout probe from the second plurality of readout probes to the second readout binding target of the one or more readout binding targets.

[0061] In some embodiments, the composition further comprises: a second overhang sequence, linked to the other end of the primary binding sequence.

[0062] In some embodiments, the composition further comprises: a second plurality of bridge probes, wherein each bridge probe comprises: a binding sequence that specifically binds to all or a part of the second overhang sequence of a primary probe of the plurality of primary probes, and one or more additional readout binding targets connected in series and linked to the binding sequence.

[0063] In some embodiments, the composition further comprises: a third plurality of readout probes, wherein each readout probe comprises: a readout binding sequence that specifically binds to a first additional readout binding target of the one or more additional readout binding targets in a bridge probe of the second plurality of bridge probes, and a signal moiety linked to the readout binding sequence via a cleavable linker.

[0064] In these embodiments, the signal moiety is capable of emitting a third detectable visual signal upon binding of each readout probe from the third plurality of readout probes to the first additional readout binding target of the one or more additional readout binding targets.

[0065] In some embodiments, the composition further comprises: a fourth plurality of readout probes. Each readout probe in the fourth plurality of readout probes comprises: a readout binding sequence that specifically binds to a second additional readout binding target of the one or more additional readout binding targets in a bridge probe of the second plurality of bridge probes, and a signal moiety linked to the readout binding sequence via a cleavable linker.

[0066] In these embodiments, the signal moiety is capable of emitting a fourth detectable visual signal upon binding of each readout probe from the fourth plurality of readout probes to the second additional readout binding target of the one or more additional readout binding targets.

[0067] In some embodiments, the cleavable linker is selected from the group consisting of an enzyme cleavable linker, a nucleophile/base sensitive linker, reduction sensitive linker, a photo-cleavable linker, an electrophile/acid sensitive linker, a metal-assisted cleavable linker, and an oxidation sensitive linker.

[0068] In some embodiments, the cleavable linker is a disulfide bond or a nucleic acid restriction site. In some embodiments, the one or more readout binding targets comprises three or more readout binding targets.

[0069] In some embodiments where second overhang is present, the additional one or more readout binding targets comprises three or more readout binding targets.

[0070] In one aspect, disclosed herein is a sequential hybridization method utilizing a plurality of primary probes, a first plurality of bridge probes, and first plurality of readout probes. In some embodiments, the method comprises the steps of: a) contacting a target nucleic acid molecule with a plurality of primary probes, where each primary probe comprises: a primary binding sequence that binds to a complementary target sequence within the target nucleic acid molecule, and a first overhang sequence connected to one end of the primary binding sequence; b) contacting, after step a) the target nucleic acid molecule with a first plurality of bridge probes, where each bridge probe comprises: a binding sequence that specifically binds to all or a part of the first overhang sequence of a primary probe of the plurality of primary probes, and one or more readout binding targets connected in series and linked to the binding sequence; and c) contacting, after step b) the target nucleic acid molecule with a first plurality of readout probes, wherein each readout probe comprises: a readout binding sequence that specifically binds to a

first readout binding target of the one or more readout binding targets of a primary probe of the plurality of primary probes, and a signal moiety linked to the readout binding sequence via a cleavable linker.

[0071] In these embodiments, the signal moiety is capable of emitting a first detectable visual signal upon binding of each readout probe from the first plurality of readout probes to the first readout binding target of the one or more readout binding targets of a bridge probe of the first plurality of bridge probes.

[0072] In some embodiments, the method further comprises the steps of: c1) imaging the target nucleic acid molecule after step c) so that interactions between the first plurality of readout probes and the first readout binding target of the one or more readout binding targets of a primary bridge probe are detected by the presence of first detectable visual signal; and c2) applying, after step c1) a cleaving agent to cleave the linker, thereby eliminating the signal moiety from each readout probe in the first plurality of readout probes.

[0073] In some embodiments, the method further comprises: d) contacting, after step c), the target nucleic acid molecule with a second plurality of readout probes. Each readout probe comprises: a readout binding sequence that specifically binds to a second readout binding target of the one or more readout binding targets of a bridge probe, and a signal moiety linked to the readout binding sequence via a cleavable linker.

[0074] In these embodiments, the signal moiety is capable of emitting a second detectable visual signal upon binding of each readout probe from the second plurality of readout probes to the second readout binding target of the one or more readout binding targets of a bridge probe of the first plurality of bridge probes.

[0075] In some embodiments, the method further comprises: d1) imaging the target nucleic acid molecule after step d) so that interactions between the second plurality of readout probes and the second readout binding target of the one or more readout binding targets of a bridge probe are detected by the presence of second detectable visual signal; and d2) applying a cleaving agent to cleave the linker, thereby eliminating the signal moiety from each readout probe in the second plurality of readout probes.

[0076] In some embodiments, each primary probe in the plurality of primary probes further comprises: a second overhang sequence connected to the other end of the primary binding sequence.

[0077] In some embodiments, the method further comprises: e) contacting, after step d), the target nucleic acid molecule with a second plurality of bridge probes. Each bridge probe comprises: a binding sequence that specifically binds to all or a part of the second overhang sequence of a primary probe of the plurality of primary probes, and one or more additional readout binding targets connected in series and linked to the binding sequence.

[0078] In some embodiments, the method further comprises: f) contacting, after step e), the target nucleic acid molecule with a third plurality of readout probes. Each readout probe comprises: a readout binding sequence that specifically binds to a first additional readout binding target of the one or more additional readout binding targets of a bridge probe in the second plurality of bridge probes, and a signal moiety linked to the readout binding sequence via a cleavable linker.

[0079] In these embodiments, the signal moiety is capable of emitting a third detectable visual signal upon binding of each readout probe from the third plurality of readout probes to the first additional readout binding target of the one or more additional readout binding targets.

[0080] In some embodiments, the method further comprises: f1) imaging the target nucleic acid molecule after step f) so that interactions between the third plurality of readout probes and the first additional readout binding target of the one or more additional readout binding targets of a bridge probe in the second plurality of bridge probes are detected by the presence of the third detectable visual signal; and f2) applying a cleaving agent to cleave the linker, thereby eliminating the signal moiety from each readout probe in the third plurality of readout probes.

[0081] In some embodiments, the method further comprises: g) contacting, after step f), the target nucleic acid molecule with a fourth plurality of readout probes. Each readout probe comprises: a readout binding sequence that specifically binds to a second additional readout binding target of the one or more additional readout binding targets of a bridge probe in the second plurality of bridge probes, and a signal moiety linked to the readout binding sequence via a cleavable linker.

[0082] In these embodiments, the signal moiety is capable of emitting a fourth detectable visual signal upon binding of each readout probe from the fourth plurality of readout probes to the second additional readout binding target of the one or more additional readout binding targets.

[0083] In some embodiments, the method further comprises: h1) imaging the target nucleic acid molecule after step g) so that interactions between the fourth plurality of readout probes and the second additional readout binding target of the one or more additional readout binding targets of a bridge probe in the second plurality of bridge probes are detected by the presence of the fourth detectable visual signal; and h2) applying a cleaving agent to cleave the linker, thereby eliminating the signal moiety from each readout probe in the fourth plurality of readout probes.

[0084] In some embodiments, the target nucleic acid molecule is an mRNA or a DNA. In some embodiments, the target nucleic acid molecule is within an intact mammalian cell. In some embodiments, the intact mammalian cell is a human cell.

[0085] In these embodiments, the cleavable linker is selected from the group consisting of an enzyme cleavable linker, a nucleophile/base sensitive linker, reduction sensitive linker, a photo-cleavable linker, an electrophile/acid sensitive linker, a metal-assisted cleavable linker, and an oxidation sensitive linker. In these embodiments, the cleavable linker is a disulfide bond or a nucleic acid restriction site. In these embodiments, the one or more readout binding targets comprises three or more readout binding targets.

[0086] In these embodiments where a second overhang is present, the additional one or more readout binding targets comprises three or more readout binding targets.

[0087] In one aspect, disclosed herein is a composition that comprises a plurality of primary probes and a first plurality of readout probes. In these embodiments, each primary probe comprises: a primary binding sequence that binds to a complementary target sequence in a target nucleic acid molecule, and a first overhang sequence connected to one end of the primary binding sequence, wherein the first overhang sequence comprises one or more readout binding targets connected in series. Also in these embodiments, each readout probe comprises: a readout binding sequence that specifically binds to a first readout binding target of the one or more readout binding targets in a first overhang sequence, and a signal moiety linked to the readout

binding sequence via a cleavable linker. In these embodiments, the signal moiety is capable of emitting a first detectable visual signal upon binding of each readout probe from the first plurality of readout probes to the first readout binding target of one of the one or more readout binding targets.

[0088] In some embodiments, the composition further comprises: a second plurality of readout probes, where each readout probe comprises: a readout binding sequence that specifically binds to a second readout binding target of the one or more readout binding targets in a first overhang sequence, and a signal moiety linked to the readout binding sequence via a cleavable linker. In these embodiments, the signal moiety is capable of emitting a second detectable visual signal upon binding of each readout probe from the second plurality of readout probes to the second readout binding target of the one or more readout binding targets.

[0089] In some embodiments, a primary probe further comprises: a second overhang sequence, linked to the other end of the primary binding sequence, where the second overhang sequence comprises one or more additional readout binding targets connected in series.

[0090] In some embodiments, the composition further comprises a third plurality of readout probes, where each readout probe comprises: a readout binding sequence that specifically binds to a first additional readout binding target of the one or more additional readout binding targets in a second overhang sequence, and a signal moiety linked to the readout binding sequence via a cleavable linker. In these embodiments, the signal moiety is capable of emitting a third detectable visual signal upon binding of each readout probe from the third plurality of readout probes to the first additional readout binding target of the one or more additional readout binding targets.

[0091] In some embodiments, the composition further comprises a fourth plurality of readout probes, where each readout probe comprises: a readout binding sequence that specifically binds to a second additional readout binding target of the one or more additional readout binding targets in a second overhang sequence, and a signal moiety linked to the readout binding sequence via a cleavable linker. In these embodiments, the signal moiety is capable of emitting a fourth detectable visual signal upon binding of each readout probe from the fourth plurality of readout probes to the second additional readout binding target of the one or more additional readout binding targets.

[0092] In any embodiments disclosed herein, the cleavable linker is selected from the group consisting of an enzyme cleavable linker, a nucleophile/base sensitive linker, reduction sensitive linker, a photo-cleavable linker, an electrophile/acid sensitive linker, a metal-assisted cleavable linker, and an oxidation sensitive linker.

[0093] In any embodiments disclosed herein, the cleavable linker is a disulfide bond or a nucleic acid restriction site.

[0094] In any embodiments disclosed herein, the one or more readout binding targets comprises three or more readout binding targets.

[0095] In embodiments where a second overhang sequence is present, the additional one or more readout binding targets comprises three or more readout binding targets.

[0096] In some embodiments, the target nucleic acid molecule is an mRNA or a DNA. In some embodiments, the target nucleic acid molecule is within an intact mammalian cell. In some embodiments, the intact mammalian cell is a human cell.

[0097] In one aspect, disclosed herein is a sequential hybridization method utilizing with a plurality of primary probes and a first plurality of readout probes. The method comprises the steps of: a) contacting a target nucleic acid molecule with a plurality of primary probes. Each primary probe comprises: a primary binding sequence that binds to a complementary target sequence within the target nucleic acid molecule, and a first overhang sequence connected to one end of the primary binding sequence, wherein the first overhang sequence comprises one or more readout binding targets connected in series; and b) contacting, after step a) the target nucleic acid molecule with a first plurality of readout probes. Each readout probe comprises: a readout binding sequence that specifically binds to a first readout binding target of the one or more readout binding targets of a primary probe of the plurality of primary probes, and a signal moiety linked to the readout binding sequence via a cleavable linker.

[0098] In these embodiments, the signal moiety is capable of emitting a first detectable visual signal upon binding of each readout probe from the first plurality of readout probes to the first readout binding target of one of the one or more readout binding targets.

[0099] In some embodiments, the method further comprises the steps of: b1) imaging the target nucleic acid molecule after step b) so that interactions between the first plurality of readout probes and the first readout binding target of the one or more readout binding targets of a

primary bridge probe are detected by the presence of the first detectable visual signal; and b2) applying a cleaving agent to cleave the linker, thereby eliminating the signal moiety from each readout probe in the first plurality of readout probes.

[00100] In some embodiments, the method further comprises the steps of: c) contacting, after step b), the target nucleic acid molecule with a second plurality of readout probes. Each readout probe comprises: a readout binding sequence that specifically binds to a second readout binding target of the one or more readout binding targets of a primary probe, and a signal moiety linked to the readout binding sequence via a cleavable linker.

[00101] In these embodiments, the signal moiety is capable of emitting a second detectable visual signal upon binding of each readout probe from the second plurality of readout probes to the second readout binding target of the one or more readout binding targets.

[00102] In some embodiments, the method further comprises the steps of: c1) imaging the target nucleic acid molecule after step c) so that interactions between the second plurality of readout probes and the second readout binding target of the one or more readout binding targets of a primary probe are detected by the presence of the second detectable visual signal; and c2) applying a cleaving agent to cleave the linker, thereby eliminating the signal moiety from each readout probe in the second plurality of readout probes.

[00103] In some embodiments, each primary probe in the plurality of primary probes further comprises: a second overhang sequence connected to the other end of the primary binding sequence, wherein the second overhang sequence comprises one or more additional readout binding targets connected in series.

[00104] In some embodiments, the method further comprises the steps of: d) contacting, after step c), the target nucleic acid molecule with a third plurality of readout probes. Each readout probe comprises: a readout binding sequence that specifically binds to a first additional readout binding target of the one or more additional readout binding targets of a primary probe, and a signal moiety linked to the readout binding sequence via a cleavable linker.

[00105] In these embodiments, the signal moiety is capable of emitting a third detectable visual signal upon binding of each readout probe from the third plurality of readout probes to the first additional readout binding target of the one or more additional readout binding targets.

[00106] In some embodiments, the method further comprises the steps of: d1) imaging the target nucleic acid molecule after step d) so that interactions between the second plurality of readout probes and the second readout binding target of the one or more readout binding targets of a primary probe are detected by the presence of the second detectable visual signal; and d2) applying a cleaving agent to cleave the linker, thereby eliminating the signal moiety from each readout probe in the second plurality of readout probes.

[00107] In some embodiments, the method further comprises the steps of: e) contacting, after step d), the target nucleic acid molecule with a fourth plurality of readout probes. Each readout probe comprises: a readout binding sequence that specifically binds to a second additional readout binding target of the one or more additional readout binding targets of a primary probe, and a signal moiety linked to the readout binding sequence via a cleavable linker,

[00108] In these embodiments, the signal moiety is capable of emitting a fourth detectable visual signal upon binding of each readout probe from the fourth plurality of readout probes to the second additional readout binding target of the one or more additional readout binding targets.

[00109] In some embodiments, the method further comprises the steps of: e1) imaging the mRNA after step d) so that interactions between the fourth plurality of readout probes and the second additional readout binding target of the one or more additional readout binding targets of a primary probe are detected by the presence of the fourth detectable visual signal; and e2) applying a cleaving agent to cleave the linker, thereby eliminating the signal moiety from each readout probe in the fourth plurality of readout probes.

[00110] In some embodiments, the target nucleic acid molecule is an mRNA or a DNA. In some embodiments, the target nucleic acid molecule is within an intact mammalian cell. In some embodiments, the intact mammalian cell is a human cell.

[00111] In some embodiments, the cleavable linker is selected from the group consisting of an enzyme cleavable linker, a nucleophile/base sensitive linker, reduction sensitive linker, a photo-cleavable linker, an electrophile/acid sensitive linker, a metal-assisted cleavable linker, and an oxidation sensitive linker. In some embodiments, the cleavable linker is a disulfide bond or a nucleic acid restriction site.

[00112] In some embodiments, the one or more readout binding targets comprises three or more readout binding targets.

[00113] In some embodiments where the second overhang sequence is present, the additional one or more readout binding targets comprises three or more readout binding targets.

[00114] In one aspect, disclosed herein is a composition comprising a first plurality of nucleic acid detection probes and an extendible signal motif formed by a first plurality populations of extender probes $\{EP_1, EP_2, \ldots, EP_n\}$. In some embodiments, each nucleic acid detection probe in the first plurality of nucleic acid detection probes comprises: a binding region comprising a binding sequence that binds to a first target sequence; and an initiator sequence linked to the binding region with a cleavable linker. In some embodiments, each population of extender probes is represented by EP_1, EP_2, ..., EP_n, respectively, where each extender probe in EP_1 comprises: a binding sequence that binds to all or a part of the initiator sequence; one or more target sequences for extender probes in EP_2 and subsequent populations of extender probes, and a signal moiety capable of emitting a first detectable signal. In some embodiments, each probe in EP_2 and subsequent populations of extender probes comprises: a binding sequence that binds to all or a part of the previous extender sequence; one or more target sequences for probes in subsequent populations of extender probes; and a signal moiety capable of emitting the first detectable signal.

[00115] In some embodiments, the first target sequence is within a primary probe that directly binds to a target nucleic acid molecule. In some embodiments, the first target sequence is within a secondary probe that binds to a primary probe that directly binds to a target nucleic acid molecule. In some embodiments, the first target sequence is within a tertiary probe that binds to a secondary probe that binds to a primary probe that directly binds to a target nucleic acid molecule.

[00116] In some embodiments, the target nucleic acid molecule is an mRNA or a DNA. In some embodiments, the target nucleic acid molecule is within an intact mammalian cell. In some embodiments, the intact mammalian cell is a human cell.

[00117] In some embodiments, the cleavable linker is selected from the group consisting of an enzyme cleavable linker, a nucleophile/base sensitive linker, reduction sensitive linker, a photo-cleavable linker, an electrophile/acid sensitive linker, a metal-assisted cleavable linker,

and an oxidation sensitive linker. In some embodiments, the cleavable linker is a disulfide bond or a nucleic acid restriction site.

[00118] In some embodiments, each extender probe of the plurality of extender probes comprises a binding sequence that is complementary to all or a part of the initiator sequence in the nucleic acid detection probe, wherein each extender probe forms a hairpin structure, and wherein the presence of the initiator sequence causes the hairpin structure to unfold and initiates a hybridization chain reaction.

[00119] In some embodiments, the composition further comprises a second plurality of nucleic acid detection probes and an extendible signal motif formed by a second plurality populations of extender probes $\{EP_{1'}, EP_{2'}, \ldots, EP_{n'}\}$. In some embodiments, each nucleic acid detection probe in the second plurality of nucleic acid detection probes comprises: a binding region comprising a binding sequence that binds to a second target sequence; and an initiator sequence linked to the binding region with a cleavable linker. In some embodiments, each population of extender probes is represented by $EP_{1'}$, $EP_{2'}$, ..., $EP_{n'}$, respectively, wherein each extender probe in $EP_{1'}$ comprises: a binding sequence that binds to all or a part of the initiator sequence; one or more target sequences for extender probes in $EP_{2'}$ and subsequent populations of extender probes; and a signal moiety capable of emitting a second detectable signal. In some embodiments, each probe in $EP_{2'}$ and subsequent populations of extender probes comprises: a binding sequence that binds to all or a part of the previous extender sequence; one or more target sequences for probes in subsequent populations of extender probes; and a signal moiety capable of emitting the second detectable signal.

[00120] In one aspect, disclosed herein is a sequential hybridization method. The method comprises the steps of: a) contacting a target nucleic acid molecule with a first plurality of nucleic acid detection probes and b) contacting, after step a) the target nucleic acid molecule with a first plurality populations of extender probes $\{EP_1, EP_2, \ldots, EP_n\}$. In some embodiments, each nucleic acid detection probe in the first plurality of nucleic acid detection probes comprises: a binding region comprising a binding sequence that binds to a first target sequence; and an initiator sequence linked to the binding region with a cleavable linker. In some embodiments, each population of extender probes is represented by EP_1, EP_2, ..., EP_n, respectively, where each extender probe in EP_1 comprises: a binding sequence that binds to all or a part of the initiator

sequence; one or more target sequences for extender probes in EP_2 and subsequent populations of extender probes; and a signal moiety capable of emitting a first detectable signal. In some embodiments, each probe in EP_2 and subsequent populations of extender probes comprises: a binding sequence that binds to all or a part of the previous extender sequence; one or more target sequences for probes in subsequent populations of extender probes; and a signal moiety capable of emitting the first detectable signal.

[00121] In some embodiments, the method further comprises: b1) imaging the target nucleic acid molecule after step b) so that interactions between the first plurality of nucleic acid detection probes and first target sequences are detected by the presence of the first detectable visual signal; and b2) applying a cleaving agent to cleave the linker, thereby eliminating the extendible signal motif.

[00122] In some embodiments, the method further comprises: c) contacting an target nucleic acid molecule with a second plurality of nucleic acid detection probes. In some embodiment, each nucleic acid detection probe in the second plurality of nucleic acid detection probes comprises: a binding region comprising a binding sequence that binds to a second target sequence; and an initiator sequence linked to the binding region with a cleavable linker.

[00123] In some embodiments, the method further comprises: d) contacting, after step c) the target nucleic acid molecule with a second plurality populations of extender probes $\{EP_{1'}, EP_{2'}, \ldots, EP_{n'}\}$, where each population of extender probes is represented by $EP_{1'}$, $EP_{2'}$, …, and $EP_{n'}$, respectively. In some embodiments, each extender probe in $EP_{1'}$ comprises: a binding sequence that binds to all or a part of the initiator sequence; one or more target sequences for extender probes in $EP_{2'}$ and subsequent populations of extender probes; and a signal moiety capable of emitting a second detectable signal. In some embodiments, each probe in $EP_{2'}$ and subsequent populations of extender probes comprises: a binding sequence that binds to all or a part of the previous extender sequence; one or more target sequences for probes in subsequent populations of extender probes; and a signal moiety capable of emitting the second detectable signal.

[00124] In some embodiments, the method further comprises: d1) imaging the target nucleic acid molecule after step d) so that interactions between the second plurality of nucleic acid detection probes and second target sequences are detected by the presence of the second

detectable visual signal; and d2) applying a cleaving agent to cleave the linker, thereby eliminating the extendible signal motif.

[00125] In some embodiments, the second target sequence is within a primary probe that directly binds to a target nucleic acid molecule. In some embodiments, the second target sequence is within a secondary probe that binds to a primary probe that directly binds to a target nucleic acid molecule. In some embodiments, the second target sequence is within a tertiary probe that binds to a secondary probe that binds to a primary probe that directly binds to a target nucleic acid molecule.

[00126] In some embodiments, the target nucleic acid molecule is an mRNA or a DNA. In some embodiments, the target nucleic acid molecule is within an intact mammalian cell. In some embodiments, the intact mammalian cell is a human cell.

[00127] In some embodiments, the cleavable linker is selected from the group consisting of an enzyme cleavable linker, a nucleophile/base sensitive linker, reduction sensitive linker, a photo-cleavable linker, an electrophile/acid sensitive linker, a metal-assisted cleavable linker, and an oxidation sensitive linker. In some embodiments, the cleavable linker is a disulfide bond or a nucleic acid restriction site.

[00128] In some embodiments, each extender probe of the plurality of extender probes comprises a binding sequence that is complementary to all or a part of the initiator sequence in the nucleic acid detection probe, where each extender probe forms a hairpin structure, and where the presence of the initiator sequence causes the hairpin structure to unfold and initiates a hybridization chain reaction.

[00129] The compositions and methods disclosed herein can be used in sequential hybridizations to identify any suitable cellular targets within an intact cell or in an *in vitro* setting. In some embodiments, the cellular targets can be mRNAs or DNAs. In some embodiments, the cellular targets can be proteins. For example, the initial target-binding primary probe can be an antibody conjugated with nucleic acid sequence for subsequent bindings.

[00130] One of skill in the art would understand that embodiments disclosed herein can be applied or combined in any aspect when applicable.

DEFINITIONS

[00131] *Animal:* As used herein, the term "animal" refers to any member of the animal kingdom. In some embodiments, "animal" refers to humans, at any stage of development. In some embodiments, "animal" refers to non-human animals, at any stage of development. In certain embodiments, the non-human animal is a mammal (*e.g.*, a rodent, a mouse, a rat, a rabbit, a monkey, a dog, a cat, a sheep, cattle, a primate, and/or a pig). In some embodiments, animals include, but are not limited to, mammals, birds, reptiles, amphibians, fish, and/or worms. In some embodiments, an animal may be a transgenic animal, a genetically-engineered animal, and/or a clone.

[00132] *Approximately:* As used herein, the terms "approximately" or "about" in reference to a number are generally taken to include numbers that fall within a range of 5%, 10%, 15%, or 20% in either direction (greater than or less than) of the number unless otherwise stated or otherwise evident from the context (except where such number would be less than 0% or exceed 100% of a possible value). In some embodiments, use of the term "about" in reference to dosages means ± 5 mg/kg/day.

[00133] *Homology*: "Homology" or "identity" or "similarity" refers to sequence similarity between two nucleic acid molecules. Homology and identity can each be determined by comparing a position in each sequence which can be aligned for purposes of comparison. When an equivalent position in the compared sequences is occupied by the same base, then the molecules are identical at that position; when the equivalent site occupied by the same or a similar nucleic acid residue (e.g., similar in steric and/or electronic nature), then the molecules can be referred to as homologous (similar) at that position. Expression as a percentage of homology/similarity or identity refers to a function of the number of identical or similar nucleic acids at positions shared by the compared sequences. A sequence which is "unrelated" or "non-homologous" shares less than 40% identity, less than 35% identity, less than 30% identity, or less than 25% identity with a sequence described herein. In comparing two sequences, the absence of residues (amino acids or nucleic acids) or presence of extra residues also decreases the identity and homology/similarity.

[00134] In some embodiments, the term "homology" describes a mathematically based comparison of sequence similarities which is used to identify genes with similar functions or

motifs. The nucleic acid sequences described herein can be used as a "query sequence" to perform a search against public databases, for example, to identify other family members, related sequences or homologs. In some embodiments, such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (1990) J. Mol. Biol. 215:403-10. In some embodiments, BLAST nucleotide searches can be performed with the NBLAST program, score=100, word length =12 to obtain nucleotide sequences homologous to nucleic acid molecules of the invention. In some embodiments, to obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) Nucleic Acids Res. 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and BLAST) can be used (See www.ncbi.nlm.nih.gov).

[00135] *Identity*: As used herein, "identity" means the percentage of identical nucleotide residues at corresponding positions in two or more sequences when the sequences are aligned to maximize sequence matching, i.e., taking into account gaps and insertions. Identity can be readily calculated by known methods, including but not limited to those described in (Computational Molecular Biology, Lesk, A. M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D. W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A. M., and Griffin, H. G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., SIAM J. Applied Math., 48: 1073 (1988). Methods to determine identity are designed to give the largest match between the sequences tested. Moreover, methods to determine identity are codified in publicly available computer programs. Computer program methods to determine identity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., Nucleic Acids Research 12(1): 387 (1984)), BLASTP, BLASTN, and FASTA (Altschul, S. F. et al., J. Molec. Biol. 215: 403-410 (1990) and Altschul et al. Nuc. Acids Res. 25: 3389-3402 (1997)). The BLAST X program is publicly available from NCBI and other sources (BLAST Manual, Altschul, S., et al., NCBI NLM NIH Bethesda, Md. 20894; Altschul, S., et al., J. Mol. Biol. 215:

403-410 (1990). The well-known Smith Waterman algorithm can also be used to determine identity.

[00136] *In vitro*: As used herein, the term "*in vitro*" refers to events that occur in an artificial environment, *e.g.*, in a test tube or reaction vessel, in cell culture, *etc.*, rather than within an organism (*e.g.,* animal, plant, and/or microbe).

[00137] *In vivo:* As used herein, the term "*in vivo*" refers to events that occur within an organism (*e.g.,* animal, plant, and/or microbe).

[00138] *Oligonucleotide:* the term "oligonucleotide" refers to a polymer or oligomer of nucleotide monomers, containing any combination of nucleobases, modified nucleobases, sugars, modified sugars, phosphate bridges, or modified bridges.

[00139] Oligonucleotides of the present invention can be of various lengths. In particular embodiments, oligonucleotides can range from about 2 to about 200 nucleotides in length. In various related embodiments, oligonucleotides, single-stranded, double-stranded, and triple-stranded, can range in length from about 4 to about 10 nucleotides, from about 10 to about 50 nucleotides, from about 20 to about 50 nucleotides, from about 15 to about 30 nucleotides, from about 20 to about 30 nucleotides in length. In some embodiments, the oligonucleotide is from about 9 to about 39 nucleotides in length. In some embodiments, the oligonucleotide is at least 4 nucleotides in length. In some embodiments, the oligonucleotide is at least 5 nucleotides in length. In some embodiments, the oligonucleotide is at least 6 nucleotides in length. In some embodiments, the oligonucleotide is at least 7 nucleotides in length. In some embodiments, the oligonucleotide is at least 8 nucleotides in length. In some embodiments, the oligonucleotide is at least 9 nucleotides in length. In some embodiments, the oligonucleotide is at least 10 nucleotides in length. In some embodiments, the oligonucleotide is at least 11 nucleotides in length. In some embodiments, the oligonucleotide is at least 12 nucleotides in length. In some embodiments, the oligonucleotide is at least 15 nucleotides in length. In some embodiments, the oligonucleotide is at least 20 nucleotides in length. In some embodiments, the oligonucleotide is at least 25 nucleotides in length. In some embodiments, the oligonucleotide is at least 30 nucleotides in length. In some embodiments, the oligonucleotide is a duplex of complementary strands of at least 18 nucleotides in length. In some embodiments, the oligonucleotide is a duplex of complementary strands of at least 21 nucleotides in length.

[00140] *Predetermined:* By predetermined is meant deliberately selected, for example as opposed to randomly occurring or achieved. A composition that may contain certain individual oligonucleotides because they happen to have been generated through a process that cannot be controlled to intentionally generate the particular oligonucleotides is not a "predetermined" composition. In some embodiments, a predetermined composition is one that can be intentionally reproduced (e.g., through repetition of a controlled process).

[00141] *Probe*: As used herein, the term "probe" or "probes" refers to any molecules, synthetic or naturally occurring, that can attach themselves directly or indirectly to a molecular target (e.g., an mRNA sample, DNA molecules, protein molecules, RNA and DNA isoform molecules, single nucleotide polymorphism molecules, and etc.). For example, a probe can include an nucleic acid molecule, an oligonucleotide, a protein (e.g., an antibody or an antigen binding sequence), or combinations thereof. For example, a protein probe may be connected with one or more nucleic acid molecules to for a probe that is a chimera. As disclosed herein, in some embodiments, a probe itself can produce a detectable signal. In some embodiments, a probe is connected, directly or indirectly via an intermediate molecule, with a signal moiety (e.g., a dye or fluorophore) that can produce a detectable signal.

[00142] *Sample*: As used herein, the term "sample" refers to a biological sample obtained or derived from a source of interest, as described herein. In some embodiments, a source of interest comprises an organism, such as an animal or human. In some embodiments, a biological sample comprises biological tissue or fluid. In some embodiments, a biological sample is or comprises bone marrow; blood; blood cells; ascites; tissue or fine needle biopsy samples; cell-containing body fluids; free floating nucleic acids; sputum; saliva; urine; cerebrospinal fluid, peritoneal fluid; pleural fluid; feces; lymph; gynecological fluids; skin swabs; vaginal swabs; oral swabs; nasal swabs; washings or lavages such as a ductal lavages or broncheoalveolar lavages; aspirates; scrapings; bone marrow specimens; tissue biopsy specimens; surgical specimens; feces, other body fluids, secretions, and/or excretions; and/or cells therefrom, *etc.* In some embodiments, a biological sample is or comprises cells obtained from an individual. In some embodiments, a sample is a "primary sample" obtained directly from a source of interest by any appropriate means. For example, in some embodiments, a primary biological sample is obtained by methods selected from the group consisting of biopsy (*e.g.*, fine needle aspiration or

tissue biopsy), surgery, collection of body fluid (*e.g.,* blood, lymph, feces *etc.*), *etc.* In some embodiments, as will be clear from context, the term "sample" refers to a preparation that is obtained by processing (e.g., by removing one or more components of and/or by adding one or more agents to) a primary sample. For example, filtering using a semi-permeable membrane. Such a "processed sample" may comprise, for example nucleic acids or proteins extracted from a sample or obtained by subjecting a primary sample to techniques such as amplification or reverse transcription of mRNA, isolation and/or purification of certain components, etc.

[00143] *Subject:* As used herein, the term "subject" or "test subject" refers to any organism to which a provided compound or composition is administered in accordance with the present invention *e.g.*, for experimental, diagnostic, prophylactic, and/or therapeutic purposes. Typical subjects include animals (*e.g.*, mammals such as mice, rats, rabbits, non-human primates, and humans; insects; worms; *etc.*) and plants. In some embodiments, a subject may be suffering from, and/or susceptible to a disease, disorder, and/or condition.

[00144] *Substantially:* As used herein, the term "substantially" refers to the qualitative condition of exhibiting total or near-total extent or degree of a characteristic or property of interest. One of ordinary skill in the biological arts will understand that biological and chemical phenomena rarely, if ever, go to completion and/or proceed to completeness or achieve or avoid an absolute result. The term "substantially" is therefore used herein to capture the potential lack of completeness inherent in many biological and/or chemical phenomena.

[00145] *Suffering from*: An individual who is "suffering from" a disease, disorder, and/or condition has been diagnosed with and/or displays one or more symptoms of a disease, disorder, and/or condition.

[00146] *Susceptible to*: An individual who is "susceptible to" a disease, disorder, and/or condition is one who has a higher risk of developing the disease, disorder, and/or condition than does a member of the general public. In some embodiments, an individual who is susceptible to a disease, disorder and/or condition may not have been diagnosed with the disease, disorder, and/or condition. In some embodiments, an individual who is susceptible to a disease, disorder, and/or condition may exhibit symptoms of the disease, disorder, and/or condition. In some embodiments, an individual who is susceptible to a disease, disorder, and/or condition may not exhibit symptoms of the disease, disorder, and/or condition. In some embodiments, an

individual who is susceptible to a disease, disorder, and/or condition will develop the disease, disorder, and/or condition. In some embodiments, an individual who is susceptible to a disease, disorder, and/or condition will not develop the disease, disorder, and/or condition.

[00147] *Treat:* As used herein, the term "treat," "treatment," or "treating" refers to any method used to partially or completely alleviate, ameliorate, relieve, inhibit, prevent, delay onset of, reduce severity of, and/or reduce incidence of one or more symptoms or features of a disease, disorder, and/or condition. Treatment may be administered to a subject who does not exhibit signs of a disease, disorder, and/or condition. In some embodiments, treatment may be administered to a subject who exhibits only early signs of the disease, disorder, and/or condition, for example for the purpose of decreasing the risk of developing pathology associated with the disease, disorder, and/or condition.

[00148] *Wild-type:* As used herein, the term "wild-type" has its art-understood meaning that refers to an entity having a structure and/or activity as found in nature in a "normal" (as contrasted with mutant, diseased, altered, etc.) state or context. Those of ordinary skill in the art will appreciate that wild type genes and polypeptides often exist in multiple different forms (e.g., alleles).

BRIEF DESCRIPTION OF THE DRAWINGS

[00149] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

[00150] Those of skill in the art will understand that the drawings, described below, are for illustrative purposes only. The drawings are not intended to limit the scope of the present teachings in any way.

[00151] Figure 1 depicts exemplary embodiments of known methods for coding molecular targets.

[00152] Figure 2 depicts exemplary sequential barcoding of provided methods. (a) Schematic of sequential barcoding. In each round of hybridization, multiple probes (*e.g.*, 24)were hybridized on each transcript, imaged and then stripped by DNase I treatment. The same probe sequences could be used in different rounds of hybridization, but probes were

coupled to different fluorophores. (b) Composite four-color FISH Data from 3 rounds of hybridizations on multiple yeast cells. Twelve genes were encoded by 2 rounds of hybridization, with the third hybridization using the same probes as hybridization 1. The boxed regions were magnified in the bottom right corner of each image. The matching spots were shown and barcodes were extracted. Spots without co-localization, without the intention to be limited by theory, could be due to nonspecific binding of probes in the cell as well as mis-hybridization. The number of each barcode were quantified to provide the abundances of the corresponding transcripts in single cells. (c) Exemplary barcodes. mRNA 1: Yellow-Blue-Yellow; mRNA 2: Green-Purple-Green; mRNA 3: Purple-Blue-Purple; and mRNA 4: Blue-Purple-Blue.

[00153] Figure 3A illustrates aspects of an exemplary hybridization experiment.

[00154] Figure 3B depicts an exemplary process for performing a pseudo-color based barcoding scheme.

[00155] Figure 4 depicts an overview illustrating a method for performing a pseudo-color based barcoding scheme: SPOTs (Sequential Probing Of Targets). (a) RNA molecules are captured on Locked Nucleic Acid(LNA) poly(dT) functionalized coverslips followed by hybridization of 28-32 gene specific primary probes. (b) Schematic to generate 12 pseudocolors used in decoding each RNA species. In each round of serial hybridizations, 3 'colors' are generated by imaging 3 unique readout probes conjugated with dye Alexa 647, Alexa 594, and Cy3b. After 4 rounds of serial hybridizations, the images are collapsed into 1 image to generate an image with 12 pseudo-colors. (c) Schematic of decoding 5 barcoded rounds based on 12 pseudo-colors coding scheme. (d) Digitized image of 12 pseudo-colors switching based on actual experimental image. After decoding 5 barcoded hyb, up to 20736 unique RNA species can be decoded with high accuracy and low error rates.

[00156] Figure 5 illustrates mRNA transcripts immobilized on a surface through poly-A-tail or hydrogel embedment.

[00157] Figure 6 illustrates an exemplary embodiment of gene specific primary probes design and sequential barcoding hybridization on mRNA immobilized on a surface. (a) Design of gene specific primary probe. (b) primary probe hybridizations on mRNA on the surface, alternatively, the primary probes hybridizations can also be done in solution. (c) sequential barcoding hybridization on the surface. One or multiple 18-30mers secondary readout probes

conjugated with fluorophore will hybridize to the primary probes during each round of hybridization.

[00158] Figure 7 illustrates different arrangements of secondary probe binding sites on primary probes. Assume 24 primary probes are used for each gene, 4 rounds of barcoding hybridization with 4 different unique secondary probe binding sequences, there are various combinations of arranging the secondary probe binding sequences on the primary probe. (a) all 4 unique binding sequences a, b, c, d are placed on all 24 primary probes. (b) each unique binding sequence are placed separately into each primary probe. In this case, 6 primary probes have unique binding sequence a, 6 primary probes will get unique binding sequence b, 6 primary probes will have unique binding sequence c, and 6 primary probes will have unique binding sequence d. (c). a combination of unique binding sequences is placed on different primary probes for a gene. In this case, the unique binding sequences combination can be (a,b),(a,c),(a,d),(b,c),(b,d),(c,d) on each 4 primary probes.

[00159] Figure 8 depicts possible ways to extinguish fluorescent signals via (a) flowing in high concentration of formamide to 'melt' off the secondary readout probe or (b) chemical cleavage.

[00160] Figure 9 depicts a schematic illustration of serial hybridization to scale up the number of available 'fluorophore'. For example, by using 3 fluorophores, and a total of 12 unique secondary readout sequences, 3 unique secondary readout probes were serially flown in. The probes code for hyb1 at a time, image, and then extinguish the signal. Then, another 3 unique secondary readout probes were be flowed in and undergo the same process. With a total of 4 rounds of serial hybridizations (n= 4), 12 unique secondary readouts with 3 colors are used to generate the barcode color for hyb1. The same process was repeated for barcoding hyb 2, … to hyb N. Essentially, a combination of serial hybridization and sequential barcoding hybridization allows us increase the number of fluorophore (F) from 3 to 12, and if number of barcoding hybridization (N) = 4, the barcode capacity will be $(3 \times 4)^4 = 20736$ which can cover almost the entire transcriptome.

[00161] Figure 10 depicts a table showing possible barcodes obtained using the pseudo-color barcode scheme.

[00162] Figure 11 depicts an exemplary embodiment of data analysis.

[00163] Figure 12 depicts an exemplary a raw image from a barcoding experiment.

[00164] Figure 13 depicts exemplary results from data analysis.

[00165] Figure 14 depicts exemplary results from data analysis.

[00166] Figure 15 depicts exemplary results from translational profiling analysis.

[00167] Figure 16 depicts exemplary results from translational profiling analysis.

[00168] Figure 17 illustrates results of gene specific primary probes design. (a) Each primary probe comprises a 25-nt gene specific sequences complementary to the mRNA, 4 readout sequences, and 2 primers binding sites. Each gene is targeted by a minimum of 28 to 32 primary probes. (b) Both priming regions (grey in the probe schematic) used in synthesizing gene specific primary probes are also used as a registration marker through the hybridization of Alexa 488 conjugated readout probes. Majority of the fluorescent spots stay even after 20 rounds of hybridizations. (Scale bars: 2μm.)

[00169] Figure 18 illustrates fluorescent switching through cleavage of disulfide conjugate dye on readout probes is highly efficient (a) 20 rounds of hybridization are accomplished by extinguishing fluorescent signals through reduction of disulfide conjugated dye to readout probes using TCEP, followed by re-hybridization of next unique secondary readout probes. The amide bond between the Alexa 488 dye (shown in yellow) and primer readout probes used as a registration marker is not affected by TCEP. (b) The fluorescent signals in each channel after treatment of 50mM of TCEP for 5 minutes at room temperature is reduced to minimal to none. (Scale bars: 5μm.)

[00170] Figure 19 illustrates RNA Sequential Probing Of Targets (SPOTs) measurement of 10,212 genes. Correlation of RNA SPOTs with RNAseq in fibroblasts with Pearson r coefficient of 0.86. (b Two SPOTs replicate experiments, signifies its high reproducibility (c) Correlation of RNA SPOTs measurement with RNAseq in mESCs, with Pearson r coefficient of 0.90. (d) Correlation of averaged smFISH counts for 24 genes in mESCs. RNA SPOTs has single molecule sensitivity and highly accurate in transcriptome profiling.

[00171] Figure 20 illustrates RNA SPOTs analysis at lower depth. (a) Correlation between RNA-seq FPKM and RNA SPOTs SPM from another replicate is high when a total of 376,781 spots are counted. SPM, spots per million; FPKM, fragments per kilobase per million reads. (b)

High reproducibility of RNA SPOTs between the two replicates in profiling ES-E14 cell gene expression (n1=376,781 spots, n2=1,688,747 spots).

[00172] Figure 21 shows raw images of 20 rounds of fluorescent switching in channel 647. Bright dots are the real targets while dim dots are due to nonspecific binding. The switching between each round of hybridization is complete, with minimal retention of fluorescent signals from previous round. (Scale bars: 2μm.)

[00173] Figure 22 depicts assessment of primary probes non-specific binding. (a) Raw images of 532 channel with the presence of mRNA on coverslips through LNA poly(d)T capturing. (b) No bright fluorescent signals is observed in the absence of mRNA on coverslips as a control. The left image has the same contrast as (a) while the right image contrast has been increased 4.5 fold to illustrate better the non-specific fluorescent signals. (c) Quantitative measurement of fluorescent intensity in channel 647 with and without the presence of mRNA. A threshold can be set to distinguish between the two populations to identify the real signals. (d) & (e) same as (c) but for channel 594 and channel 532.

[00174] Figure 23 depicts smFISH measurement in single cells correlates with RNA SPOTs measurement in NIH/3T3 cells. (a) Raw images of the 7 genes measured by smFISH in NIH/3T3 cells. (Scale bars: 5μm.) (b) The averaged RNA smFISH counts agrees with RNA SPOTs SPM (spots per million) with a Pearson correlation coefficient of 0.88, indicating RNA SPOTs quantitation is accurate. Error bars represents the standard error of the mean (SEM) across different single cells.

[00175] Figure 24 depicts that mRNA can be immobilized by polyacrylamide hydrogel on a bind-silane treated coverslips. (a) mRNA is trapped in the hydrogel mesh once acrylamide and bis-acrylamide monomers crosslink completely on the coverslip. (b) smFISH detection of ACTB once the total RNA is captured on a coverslip through LNA poly(d)T capturing (left) or polyacrylamide hydrogel (right). Negative control (channel 488) shows that the fluorescent signals are not coming from nonspecific sources. (Scale bars: 5μm.)

[00176] Figure 25 depicts exemplary reaction scheme for synthesizing DNA probes conjugated to dye through cleavable disulfide linker.

[00177] Figure 26. An exemplary embodiment illustrating sequential barcoding using gene specific primary probes, secondary bridge probes and tertiary readout probes. (a) Sequential barcoding FISH (seqFISH) with DNA readout probes conjugated with dyes through disulfide linkage. The scheme begins with hybridization of gene specific primary probes, followed by secondary bridges with readout binding sites, and a unique tertiary readout probes with disulfide-linked dye. Once imaged, reducing agent such as TCEP/DTT is used to eliminate the fluorescent signals. Subsequent hybridization gives fluorescent signals which is not interfered by previous fluorescent spots. The secondary bridges can be stripped off by using high concentration of formamide, and replaced by a new set of secondary bridges. (b) An exemplary embodiment illustrating primary probes with two overhang sequences. One of the alternate designs of gene specific primary probes and secondary bridges. For example, with 2 overhangs on primary probes, each overhang can bind 1 secondary bridge which consists of 3 unique tertiary readout probes binding sites. By using 4 different colors of fluorophore, one can scale up the barcodes to $46 = 4096$ with this design.

[00178] Figure 27A depicts an exemplary hybridization chain reaction (HCR).

[00179] Figure 27B depicts an exemplary readout probe.

[00180] Figure 27C depicts an exemplary hybridization chain reaction based on readout probes with cleavable linkers.

[00181] Figure 28 depicts an exemplary embodiment illustrating rehybridization in mouse embryonic stem cells (mESCs). (a) First hybridization with tertiary readout probes conjugated with A647. Real fluorescent spots are shown in red dashed box. (b) No fluorescent spots are observed in channel 594 during the first hybridization. (c) After TCEP treatment and washing steps, channel 647 is reimaged. The observed dim dots are non-specific sticking of dyes which does not interfere with subsequent real fluorescent spots identification. (d) Second unique readout probes are hybridized to the secondary bridge binding sites to give real fluorescent spots that appear as the same positions in (a).

[00182] We also capture in vitro transcribed polyA-tailed dCAS9-EGFP mRNA on a dT20 Locked Nucleic Acid (LNA) surface-modified coverslips to show the rehybridization scheme works on the coverslips.

[00183] Figure 29 depicts an exemplary embodiments illustrating rehybridization on mRNA captured on a dT20 LNA surface-modified coverslips. (a) (Left) First round of hybridization with tertiary probes conjugated with A647. (Right) No fluorescent spots are observed in channel 594 during first hybridization. (b) (Left) Channel 647 has minimal to no leftover fluorescent signals. (Right) Fluorescent spots from second round of hybridization which appear as the same positions as first hybridization.

[00184] Figure 30 illustrates an exemplary process for error correction.

[00185] Figure 31 illustrate an exemplary computer system for implementing the error correction methods disclosed herein.

[00186] Figure 32 depicts an overview of the Sequential barcode FISH (seqFISH) in brain slices. A). A coronal section from a mouse brain was mounted on a slide and imaged in all boxed areas. Each image was taken at 60x magnification. B). Example of barcoding hybridizations from one cell in field from A. The same points are re-probed through a sequence of 4 hybridizations (numbered). The sequence of colors at a given location provides a barcode readout for that mRNA ("barcode composite"). These barcodes are identified through referencing a lookup table abbreviated in D and quantified to obtain single cell expression. In principle, the maximum number of transcripts that can be identified with this approach scales to FN, where F is the number of fluorophores and N is the number of hybridizations. Error correction adds another round of hybridization. C). Serial smHCR is an alternative detection method where 5 genes are quantified in each hybridization and repeated N times. Serial hybridization scales as F*N. D). Schematic for multiplexing 125 genes in single cells. 100 genes are multiplexed in 4 hybridizations by seqFISH barcoding. This barcode scheme is tolerant to loss of any round of hybridization in the experiment. 25 genes are serially hybridized 5 genes at a time by 5 rounds of hybridization. Each number represents a color channel in single molecule HCR. As a control, 5 genes are measured both by double rounds of smHCR as well as barcoding in the same cell. E. SmHCR amplifies signal from individual mRNAs. After imaging, DNAse strips the smHCR probes from the mRNA, enabling rehybridization on the same mRNA (step a). The "color" of an mRNA can be modulated by hybridizing probes that trigger HCR polymers labeled with different dyes (step b). mRNA are amplified following

hybridization by adding the complementary hairpin pair (step c). The DNAse smHCR cycle is repeated on the same mRNAs to construct a predefined barcode over time.

[00187] Figure 33 illustrates an example accurate in situ quantification of mRNA levels generated by seqFISH. A). Image of seqFISH barcoding 100 genes in the outer layer of the mouse cortex. RNA dots in the image are z projected over 15μm. Individual mRNA points are shown across 4 hybridizations in the inset images. White squares correspond to identified barcodes, yellow squares correspond to missing transcripts in a particular hybridization, red squares correspond to spurious false positives and are not counted in any barcode measurements. Numbers in the squares correspond to barcode indices. B). seqFISH correlates with smHCR counts. After barcoding, 5 target mRNAs were measured twice by smHCR in the same cells, providing absolute counts of the transcripts. The two techniques correlate with an R=0.85 and a slope (m) of 0.84 (n=3851 measurements). The 2D histogram intensity shows the distribution of points around the regression line. A high density of points is seen along the regression line. The density falls off steeply around the regression line. C). Error correction results in a median gain of 373 (25%) counts per cell (n=3497). Red and blue curves correspond to the total barcode counts per cell before and after error correction. D). Dropped and off-target barcodes represent a small source of error in seqFISH. 100 on-target barcodes and 525 off-target barcodes are measured per cell. Dropped barcodes are due to at least two overlapping dots appearing within the same region. E. Off-target barcodes are rarely observed and contribute minimally to the expression profile in single cells. Each of the 100 on-target barcodes (blue) and 525 off-target barcodes (red) are quantified per cell. The mean is shown with shaded regions corresponding to 1 SD (N=41 imaged regions).

[00188] Figure 34 depicts an example illustrating that distinct clusters of cells exhibit different regional localization in the brain. A). Gene expression of 14,908 cells presented as a Z-score normalized heatmap. B). Regional compositions of 13 cell clusters are visualized as stacked bar plots with the area corresponding the number of cells in each region. Hippocampal regions are: CA3, CA1, Dentate Gyrus (DG). Cortical regions: parietal and temporal. Box plot of the Z scores of 21 representative genes are plotted for each cell class. The major tick marks correspond to Z score 0 while every minor tick is a z score interval of 1. Cell type assignments are shown on the dendrogram. Abbreviations: Hippocampus pyramidal (Hipp), cortex (Cort),

Dentate Gyrus (DG), Interneurons (Int), Astrocyes(Astro), Microglia (μGlia). C). Subclusters of cluster 6 cells and their regional localization and gene expression profile displayed under the dendrogram. Subcluster 6.1 is enriched in the CA3, while 6.7 is enriched in the DG. D). Subclusters of cluster 7 cells are shown. Almost all cells are localized in the GCL but have different combinatorial expression profiles. Note Calb1 expression, which marks out granule cell maturation, differs amongst subclusters. E). Any random subset of 25 genes can recapitulate approximately 50% of the information in the correlation amongst cells (red), but a larger number of genes are required to accurately assign cells to cluster using a random forest algorithm (blue) (n=10 bootstrap replicates; shading is 95% CI), indicating that fine structures in the data require quantitative measurements of combinatorial expression of many genes. F). Similar to E, while the first ten PCs explain the coarse structure, a larger number of principal components (PCs) are required to describe the full data. Expected variation (green) and accuracy in predicting cell identity using a random forest model (blue).

[00189] Figure 35 depicts an example embodiment, illustrating spatial layering of cell classes in the Dentate Gyrus (DG). A-B). Suprapyramidal and infrapyramidal blades of DG. Cells of the subgranular zone (SGZ) and granule cell layer (GCL) are arranged in lamina layers in mirror symmetric patterns on the upper and lower blades. C). The SGZ stays on the inner layer of the DG fork. D). Cells are patterned in the crest. Numbered color key corresponds to cluster numbers in Figure 34b. E). Letters in the cartoon of DG correspond to images. F). 3D image of the fork region shown in C).

[00190] Figure 36 depicts an example embodiment, illustrating that subregions of the hippocampus are composed of distinct compositions of cell classes based on the first 125 gene experiment. Upper right panel. Cartoon of hippocampus with imaged regions labeled. Color key corresponds to the classes in Figure 39b. A-D). These images are regions from the CA1d. Astrocytes (Astro) are marked in image A) and a microglia cell (μGlia) is marked in image B). Moving along the hippocampus from CA1 dorsal to ventral, cell classes transition from a homogenous dorsal population (C to D) to a mixed population in the CA1 intermediate (E-F) to regions of even larger cellular diversity in the CA1 ventral region (G-I). The dotted line in D) marks the transition point of the CA1d to the CA1i. E) shows two laterally segregated cell classes (marked by a dotted line) in the CA1i along with cholinergic interneurons (Int) on the

interior surface of the CA1i. The ventral (J-K) and intermediate CA3 (L-M) have similar cell classes compositions to the CA1v and CA1i. The two last regions (O-P) of the dorsal CA3 shows distinct cell classes compositions that are relatively homogeneous within a field but are different than other fields of CA3. The cell class composition of field P is similar to that of the CA1d, but these cluster 6 cells are grouped into a distinct subcluster.

[00191] Figure 37 depicts an example embodiment, showing mapping of cell types to a second brain slice with 125 genes. Upper right panel. Cartoon of hippocampus with imaged regions labeled. Color key corresponds to the classes in Figure 34b. A-D. Similar to the cell class compositions shown for the hippocampus in Figure 36, CA1d in this second coronal section from a second mouse is composed of mostly cluster 6 cells. (E) CA1i region and (F-G) the CA1 ventral regions are again composed of similar cell classes to that shown in Figure 36 with increasing diversity of cell class compositions from the CA1d to the CA1i to finally the CA1v. (H-J) CA3 regions. (K-M) DG regions showing the same cell classes and layer pattern of the GCL and SGZ shown in Figure 35.

[00192] Figure 38 depicts an example embodiments, showing mapping of cell types to a third brain slice with 249 genes. Upper right panel. Cartoon of hippocampus with imaged regions labeled. Color key corresponds to the classes in Figure 46C. A-C). Similar to the slice shown in Figures 36 and 37, CA1d is relatively homogenous in cell cluster composition. D-G). Images from the CA1i region show that the cell class composition is different from that of the CA1d. H-K). Again, similar to Figures 36 and 37, images from the CA1 ventral regions shows a much more complicated cellular composition and a high degree of cellular heterogeneity. L-R). Images from the CA3 region show that the cellular compositions also creates 3-4 subregions within the CA3. The cellular heterogeneity of the CA3 subregions mirrors that of the CA1, where the ventral region of the CA3 is very heterogenous while the dorsal region of the CA3 is relatively homogenous. S-T). The DG regions show the distinct SGZ versus GCL layering pattern seen in the previous two brains.

[00193] Figure 39 depicts an example embodiment, showing correlations of the transcription profile across the pyramidal layer A). mRNA counts in the cell bodies in the Stratum Pyramidale (SP) are grouped within each field of view. A single cell in the Stratum Radiatum (SR) is shown to illustrate individual mRNA localization. Stratum Oriens (SO) is

labeled for orientation. B). mRNAs in different subregions of pyramidal layer show both long-distance spatial correlations as well as local correlations between neighboring fields. Both CA1 and Dentate Gyrus (DG) show high regional correlations. Correlation is calculated based on the 125 gene experiment. C). Illustration of regional and long distance correlation patterns observed in B. Correlated regions are colored and long distance correlations are shown as dotted lines with their median correlation coefficient written over the dotted line.

[00194] Figure 40 depicts an example embodiment, showing barcode assignments for all genes in the combined hybridization experiment (Figure 32). Barcode assignments in the 125-gene seqFISH and serial experiment (Figure 32). 125 genes are profiled, 100 of which are barcoded and 25 are identified by serial smHCR hybridizations. Five control genes (Hdx, Vps13c, Zfp715, Fbll1, Slc4a8) were quantified by both techniques. The smHCR round of hybridization of control genes were performed twice to co-localize signal to obtain an absolute count.

[00195] Figure 41 depicts an example embodiment, showing smHCR performance metrics as compared to smFISH, (related to Figure 32). A). Raw data of Pgk1 transcripts imaged in a brain slice. The transcript was targeted with 2 hcr probes sets and 1 smFISH probe set, each consisted of 24 oligonucleotide probes. The probe sets were hybridized together and were imaged in 3 different channels. Green circles are transcripts detected in all channels, yellow circles signify transcripts detected in 2 out of 3 channels, and red circles represent signal found in only 1 channel (false positives due to nonspecific binding). T hese images show that smHCR and smFISH have similar sensitivity, specificity, and spot size. B). Gain of smHCR vs smFISH. The mean gain of smHCR is 22.1 ± 11.55 vs smFISH (n=1338). C). True positive detection rate of smHCR and smFISH per channel. The percent of true positives (transcripts detected with at least 2 out of 3 probe sets) detected with each probe set (n=1338). D). False positive rate of smHCR and smFISH. Percent of total dots in a channel not detected in any other channel for 3 color Pgk1 (n=1338). E). All the regions imaged in the coronal section are boxed. Each box represents a field of 216 um x 216 um. The brain section used for Figures 32 and 33 is shown on the left. The middle section is used for Figure 34 and the right section is used for Figure 38.

[00196] Figure 42 depicts an example embodiment, showing quantitation of seqFISH (related to Figure 33). A). All control genes show high correlations between seqFISH and

smHCR. B). Number of dropped hybridizations from the barcode. Blue bars represent measured probability and the red bars represent inferred values from binomial distribution fitting of measured probability. The ratio of the full barcodes (4 hybridizations) vs 3 hybridization barcodes indicate that transcripts that are mis-hybridized in 2 rounds are rare. Transcripts missed in 2 or more hybridizations (red bars) could not be recovered from the error-correction algorithm and would be dropped from our quantifications (N=2,115,477 total barcodes). C). Intensity of barcode hybridizations overtime. All dots belonging to barcodes are quantified in each hybridization and their mean intensity is plotted over time normalized to the first hybridization. 99% CI ratio of mean is plotted as a bar over points, but is not visible due to its small size (n=60143 to 111284 points per channel). D). Barcoding confidence ratio. Barcode classes in D) are compared to a null model of barcode observations where random chance observation should give a ratio of 1. Off target barcodes are observed 0.005 times less than expected, suggesting that seqFISH has high accuracy in correctly counting barcoded transcripts (n=3493 cells). Dark bars on top of bar plots correspond to 99.999% confidence interval determined by bootstrap resampling. E). Comparison of average copy numbers per gene as measured by Zeisel et al. and seqFISH. Single cell RNA-seq underestimates copy numbers compared to seqFISH.

[00197] Figure 43 depicts an example embodiment, showing gene expression patterns and clustering of the 125-gene dataset (related to Figure 34). A). Overview of 125 gene expression. Plots show the distribution of each transcript in all 14,908 imaged cells. Note the last 25 genes have higher expression and were imaged with serial hybridization. B). Violin plots of Z-score distribution for 125 genes. C). Sub-cluster hierarchy of each of the 13 clusters identified in Figure 34B. D). PCA eigenvalue analysis of the cell-to-cell correlation matrix. First 125 PC and their eigenvalues are shown. As observed in Figure 31, the first 10 PCs explain 59.5% of the variation in the data, while the remaining 115 PCs are needed to explain remaining data. Reflecting this, the eigenvalues of the first 10 components are high, while the remaining eigenvalues are uniform. E). Correlation between gene expression and spatial localization. Each dot represents a pair of cell classes and their correlations in gene expression space (x) and spatial localization patterns (y) (N=153 pairwise correlations between classes, R=0.67). Classes that are similar in expression have similar localization patterns. F). PCA decomposition separates cells

into coherent clusters corresponding to cell classes. Cells are colored according to the clusters displayed in the dendrogram.

[00198] Figure 44 depicts an example embodiment, showing robustness of cell classes to downsampling of cells (related to Figure 34). To measure how well cluster assignments perform with a limited number of cells, a random forest model was trained on the cell-to-cell correlation matrix of the 6872 cells in the center field of view. The robustness of the clusters was calculated by applying this model to classify the remaining cells and determining the percent accuracy of correct assignment to the clusters presented in Figure 34b. While some classes can be assigned accurately even with a small number of cells as the initial training set, several classes require large number of cells to accurately assign (n=10 bootstrap replicates, S.E.)

[00199] Figure 45 depicts an example embodiment, showing cell-to-cell correlation analysis as a function of dropping genes (related to Figure 34). A). Clustered gene to gene correlation map for all 125 genes. There are many blocks of highly correlated genes. A few genes do not fall into any blocks. B). The full cell-to-cell correlation map using all genes in the data set. C). Representative cell-to-cell correlation with the indicated number of genes used to construct the matrix indicated above each plot. Dropping genes from the data results in degradation of the fine structure of the correlation map.

[00200] Figure 46 depicts an example embodiment, showing gene expression patterns and clustering of the 249-gene dataset (related to Figure 38). A). Overview of 249-gene expression. Plots show the distribution of each transcript in all 2050 imaged cells in the hippocampus. Note the last 35 genes have higher expression and were imaged with serial hybridization. B). Violin plots of Z-score distribution for 249 genes. C). Dendogram with regional localization of the 18 cell clusters for the 249-gene experiment. D). Correlation of seqFISH counts to smHCR counts for the 249-gene experiment. The 2D density histogram shows a high density of points around the regression line that fall off towards the edges of the distribution. E). Cell-to-cell correlation for all 2050 cells in the 249-gene dataset. F). Heat map of the percentage of each cell class in each region of the hippocampus for both the 125-gene experiments. These heat maps show that in both 125-gene experiments the same cell classes are used in roughly the same proportions in each subregion. G). Heat map of the percentage of each cell class in each region of the

hippocampus for the 249-gene experiment. The same patterns are seen as the 125 gene experiment (i.e., different regions use different cell classes in varying amounts).

[00201] Figure 47 depicts an example embodiment, showing marker genes expression in the hippocampus (related to Figure 38). A). The top panel outlines the region of the hippocampus being shown in a yellow box. The images show the raw gene expression patterns seen using smHCR in our data at the dorsal most tip of the CA3 for a representative set of cell identity markers used in the 249 gene experiment. The transcript expression profile is shown in red, Nissl staining is shown in green, and DAPI staining is shown in blue. Each image shown is the full field of view and a maximum intensity projection over 15 um. B). Set of images showing the distinction between the GCL and SGZ. The GCL shows a high level of Nissl staining and expression of neuronal genes such as slc17a7 and camkII. The SGZ shows an absence of Nissl staining and terminal neuron marker genes. The transcript expression profile is shown in red, Nissl staining is shown in green, and DAPI staining is shown in blue. Each image shown is the full field of view (216 um x 216 um) and a maximum intensity projection over 15 um.

[00202] Figure 48 depicts an example embodiment, showing comparison of SeqFISH expression data to Allen Brain Atlas expression data (related to Figure 39). A). ISH data from the Allen Brain Atlas for genes seen to be enriched in the SGZ in the 125 and 249 gene seqFISH experiments. In the 125 gene experiment, mertk and mfge8 were found to be enriched in the SGZ. In the 249 gene experiment, nfia and sox11 were seen to be enriched in the SGZ. ABA ISH data shows similar patterns to those observed with seqFISH for the SGZ. B-C). Comparison of averaged z-score values per cell from seqFISH to ABA data across hippocampus. B). Amigo2 Z-score profile found across the different fields of the hippocampus using seqFISH is shown on top and the ABA ISH image for Amigo2 is shown on the bottom. C). Gpc4 Z-score profile found across the different fields of the hippocampus using seqFISH is shown on top and ABA ISH image for Gpc4 is shown on the bottom.

DETAILED DESCRIPTION OF CERTAIN EMBODIMENTS

[00203] Among other things, the present invention provides new methods, compositions and/or kits for profiling nucleic acids (*e.g.*, transcripts and/or DNA loci) in cells.

[00204] In some embodiments, the present invention provides methods for profiling nucleic acids (e.g., transcripts and/or DNA loci) in cells. In some embodiments, provide methods profile multiple targets in single cells. Provided methods can, among other things, profile a large number of targets (transcripts, DNA loci or combinations thereof), with a limited number of detectable labels through sequential barcoding.

Pseudo-Color Based Barcoding

[00205] In one aspect, disclosed herein are pseudo-color based barcoding method. For example, pre-designed barcodes are associated specific molecular targets through sequential hybridization experiments. A pseudo-color based barcoding scheme is developed to overcome limitations in the previous generation of the technology such as lack of visual signals that can be associated with the probes or small internal within cell when carrying out in situ experiments.. The current method can be applied to both in vitro and in situ analysis. According to the method, each barcoding round comprises multiple serial hybridizations where a small number of colored signals (that are associated with probes) are used in each hybridization experiment within a serial hybridization round. Images from each serial hybridization experiment within the same serial hybridization round are combined to form a composite image for each barcoding round. In each barcoding round, the same set of molecular targets are analyzed. After all barcoding rounds are completed, associated of the barcode with these molecular targets is completed.

[00206] To distinguish from existing FISH methods, the currently pseudo-color based barcoding methods is referred to as Sequential Probing Of Targets ("SPOTs").

[00207] SPOTs offers numerous advantageous over existing barcoding methods. For example, it does not require a user to have a large number of detectable label molecules, which can save time and money.

[00208] Pseudo-color scheme can overcome this density problem in both in situ and in vitro applications. In the in vitro cases, implemented with SPOTS, capturing transcripts onto an oligonucleotide dT surface and adjusting the dilution factors can easily remove the optical crowding problems and allow the transcriptome to be imaged by seqFISH.

[00209] In addition, pseudo-color is more efficient in terms of imaging time than all existing imaging methods including expansion microscopy. For example, with 3 fluorophores, it

takes 3^(10-1)=19,683 to code for the transcriptome with one round of error correction. Thus, a total of 30 frames of imaging is required. In the pseudo color scheme, 20 pseudo-colors can be used for 4 rounds of hybridization to code for 20^(4-1)=8000 genes in each of the three fluorophore channels for a total of 24,000 genes. This requires 3x20x4=240 frames to image, a 8 fold increase in the imaging time. However, because the density of mRNA is effectively diluted into 3x20=60 pseudo-channels instead of 3 fluorophore without pseudo-color coding, the density problem is alleviated with a factor of 20. The target spots can be localized to nanometer precision by Gaussian fitting to decrease the density in each pseudo-color channel before barcode alignment. Thus, the benefit of pseudo-color coding is to decrease the density of the target spots in the cell, while saving imaging time compared to expansion microscopy, where expanding the sample by 20 folds requires an additional 20 fold increase in imaging time. Since imaging time is rate limiting in any sequential imaging method, pseudo-color solves a major problem in implementing transcriptome profiling in situ.

Overview

[00210] In fluorescence in situ hybridization (FISH) methods, fluorescence labels that are capable of producing different visual signals are used to detect multiple molecular targets (such as mRNA transcripts) at the same time in one hybridization experiment.

[00211] Figure 1 illustrates three types of hybridization methods, where combinations of visual signals (barcoding schemes) are used to specifically encode individual molecular targets. Figure 1a) shows a spatial barcode scheme where different colored probes are bound to the same target. The probes of the same color are grouped and spatially separated from another color.

[00212] Figure 1b) depicts a spectral barcode scheme where different colored probes are no longer separated in groups according to their respective colors. Probes bound along a target are resolved by spectral methods to create a barcode.

[00213] The approaches in Figures 1a) and 1b) both rely on combinations of signals (colors) and are heavily limited by the types of signals that are available and the resolution at which microscopic instruments can resolve such signals.

[00214] As noted in Figure 1, when 5 different color dyes are associated with probes to identify different molecular targets, the spatial barcoding scheme in Figure 1a) of Figure 1 will

lead a barcoding capacity of 720; i.e., up to 720 molecular targets can be uniquely coded. Using the approach of Figure 1b), only a maximum of 31 molecular targets can be uniquely coded.

[00215] The approach in Figure 1c), known seqFISH, greatly expanded coding capacity. Here, instead of cramming all coding signals into a single hybridization experiment, barcodes are formed via temporally separately sequential hybridization experiments. The color of probes binding to a target can remain the same or change to a different color in different rounds. The barcode for a target is determined by the order by which each color appears during sequential rounds of hybridization. Details of this barcoding scheme are depicted in Figure 2 and can be found in International Patent Publication No. WO 2014/182528, which hereby incorporate by reference in its entirely.

[00216] This method involves lysing cells and immobilized the mRNA on a surface for single molecule quantification. However, this technology is not limited to profiling only the mRNA species, but also applicable to other molecules such as DNA and proteins.

[00217] Since mRNAs are immobilized on a surface, they can be barcoded by introducing several rounds of hybridizations, imaging the fluorescent spots, and extinguishing the signal for the next round of hybridization. The number of barcodes available scales as FN where F is the number of fluorophores and N is the number of barcoding hybridization rounds. For example, with 3 fluorophores and 9 rounds of hybridizations, the available barcodes are 3^9=19683. This method achieves single molecule sensitivity which allows the determination of gene expression through counting each identified mRNA molecules on the surface.

[00218] As shown in Figure 2a), in each round of hybridization, multiple probes (e.g., 20-30) were hybridized to each transcript, imaged and then stripped by DNase I treatment before another set of probes are used in the next round of hybridization. Probes with different binding sequences can be used in different rounds of hybridization. Also, the same probe sequences could be used in different rounds of hybridization, but probes were coupled to different fluorophores. A barcode for this mRNA is formed by combining the fluorescent colors associated with the probes in each round of hybridization according to the order or sequence by which the colors are detected. In the example illustrated in Figure 2a, the barcode for the mRNA in N rounds of hybridization will be {purple-blue-...-green}, with each color representing a particular hybridization round.

[00219] Figure 2b) illustrates exemplary composite four-color FISH images from 3 rounds of hybridizations on multiple yeast cells. As illustration, the same small region is magnified to reveal colored spots representing binding to probes to mRNA transcripts. Colored spots corresponding to the same target from different rounds of hybridization are matched to render a color barcode for the target. In each round of hybridization, the same spots were detected, but the dye associated with the transcript can change. The identity of an mRNA is encoded in the temporal sequence of dyes hybridized. For example, three rounds of hybridizations created the barcodes in Figure 1c): mRNA 1: Yellow-Blue-Yellow; mRNA 2: Green-Purple-Green; mRNA 3: Purple-Blue-Purple; and mRNA 4: Blue-Purple-Blue.

[00220] As noted above, the method of Figures 1c and 2 has significantly increased coding capacity. Based on 5 different visual signals and 7 rounds of hybridization experiments, a total of 78,125 targets (e.g., genes or mRNA transcripts) can be unique coded.

[00221] However, the types of color signals that can be used in these sequential hybridization experiments are limited. In order to encode the human genome, which includes about 20,000 protein encoding genes, at least 7 rounds of hybridization experiments are required. This many number of hybridization rounds result in long and complex barcodes that are hard to distinguish from each other, thus resulting in errors and inaccuracies. If one would like to include error correction in the barcodes, at least 8 rounds of hybridization experiments are needed, leading to even more complex barcodes and likelihood of more errors and inaccuracies.

[00222] Efforts made for creating more types of visual signals have not led to significant improvement. More improvements are need to increase coding capacities, enhance detection accuracies and sensitivities, and reduce errors.

[00223] In one aspect, a novel barcoding scheme is provided. Here, instead of developing different colors that are suitable for probe design, a counter-intuitive approach is adopted to create pseudo-colors to expand capacity of barcoding. Although the term "pseudo-color" is used, one is not limited to using colors in this new coding scheme, symbols, letters, numbers, 2D barcodes, 3D barcodes, and combinations thereof can be used to uniquely identify molecular targets. As disclosed herein the number of pseudo-colors far exceeds the number of actual colors that are associated with the detection probes used in the hybridization experiments.

[00224] Figure 3A illustrates exemplary aspects that may contribute to error correction during a sequential hybridization process. Such aspects include but not limited to sample processing 302, probe design, barcode design, hybridization, image collection, signal removal and re-hybridization, and data analysis. In practice, any or all of these aspect can contribute to the quality of analysis.

[00225] Barcoding by sequential hybridization includes multiple rounds of hybridization. Each round of hybridization in turn is a multiple step process including most or all of the aspects outline above. Errors and inaccuracies can be introduced at any step during any round of hybridization. Such errors can lead to misidentification of target genes in a sample.

[00226] Prior to hybridization, samples that will be subject to analysis are processed. The main purpose of such processing is to immobilize target molecules; for example, mRNAs, chromosomal DNAs, and proteins. It is essential that the target molecules remain spatially fixed through different rounds of hybridization.

[00227] Probe design contributes to specificity of binding between the probes and target sequences. It is possible to apply hybridization chain reaction to allow multiple probes to bind at the same target sequence to amplify detectable signals. Additionally, as illustrated in Figures 26 and 27, it is possible to insert a cleavable linker between the binding sequence (that binds a target sequence) and signal moiety (that emits visible signals) of a probe. Here, error can be reduced because no removal of probes is needed for the next round of hybridization. Instead, only visible signals are switched.

[00228] Barcodes implemented during the analysis are unique. Nonspecific binding or other mistakes can render the results from one or more rounds of hybridization unreliable. A simple solution is to remove data that are unreliable. However, if data from one or more rounds of hybridization are eliminated from analysis, some of the barcodes would become indistinguishable from each other.

[00229] During and after hybridization of probes to target sequences, there are also aspects that are important for improving the quality of the sequential analysis. For example, the hybridization conditions should be designed to avoid non-specific bindings. This can also be achieved through sample processing and probe design. Similarly, image collection can also be affected by a number of factors including sample processing, probe design and barcode design.

As described above, probes with too many types of color signal in many rounds of hybridization can lead to barcodes that are hard to resolve or even errors.

[00230] Between hybridization rounds, old probes can be removed before new probes are added. Here, the removal process can also be associated with errors. For example, if the removal condition is too harsh, immobilized biological samples can be disturbed. As a result, positions of visual signals in images from different rounds of hybridization experiments would change.

[00231] Some of the errors may be corrected or reduced by data analysis. For example, in most scenario, bindings between probes and target sequences are observed as colored bright spots over relatively darker background where no binding is observed. The spots are brightest in the center and fade away at the edges. Gaussian distribution analysis can be used to focus on the most significant portion of an image thus leading to better resolved image data. In addition, noise reduction can be used to reduce background signals.

[00232] Figure 3B depicts an exemplary process for performing a pseudo-color based barcoding scheme. Exemplary embodiment 300 provides an example of a pseudo-color based barcoding process. Here, instead of using a large number of actual detectable color signals, which are not available or impractical to obtain, a small number of color probes are repeatedly used in multiple serial hybridization experiments where non-overlapping sets of molecular targets are analyzed. Images from the serial hybridization experiment are combined to form a composite image representing a single barcoding round. In the composite image, at least some of the would-be redundant colors are replaced with predefined symbols in a pseudo-color based scheme. The process is repeated multiple times to produce the final multi-component barcodes.

[00233] At step 302 liquid molecular targets in a biological sample that needs to be barcoded will be identified. In particular, the number of the molecular targets and their respective identity, such as the name of the target and sequence information of the molecular target, will be identified to provide information that will be used for specific binding probe design.

[00234] Prior to actual hybridization experiments, the molecular targets in the biological sample will be immobilized; for example, on a glass cover slip. As disclosed herein, exemplary

biological sample includes but is not limited to a tissue sample, a cell sample, a cell extract sample, a nucleic acid sample, a RNA transcript sample, a protein sample, or an mRNA sample.

[00235] At step 304, depending on the total number of the biological molecular targets, the number of barcoding rounds will be decided. Because the number of barcoding rounds corresponds to the size of the barcode that will result, the user should decide on how many rounds will be needed depending on the sample size of the molecular targets. In some embodiments, a user may take into consideration the types of color probes that are available. In some embodiment, three or more barcoding rounds will be used. In some embodiments, four or more barcoding rounds will be used. In some embodiments, five or more barcoding rounds will be used. In some embodiments, six or more barcoding rounds will be used. As disclosed air in one advantage of The method disclosed herein is providing relatively simpler barcodes that are more error resistant. In general, a five component barcode is more preferred then an eight component barcode.

[00236] At step 306, the number of symbols in a pseudo-color scheme will be determined based on the total number of the molecular targets (e.g., $N \gg 2$) and the intended for number of barcoding rounds (e.g., $n \geq 2$). As disclosed here in the minimum number of pseudo-color symbols (e.g., S) can be determined according to the following equation:

$$S \geq \sqrt[n]{N}, \text{ where S, n and N are all integers.} \quad (1)$$

[00237] In some embodiments, one or more error correction coding rounds (e.g., x rounds where $x \geq 1$) will be implemented. As such, one may choose a higher number of pseudo-color symbols to avoid too many barcoding rounds. The minimum number of pseudo-color symbols, when one or more error reaction rounds, can be determined according to the following equation:

$$S \geq \sqrt[n-x]{N}, \text{ where S, n, x, and N are all integers.} \quad (2)$$

[00238] Although the term pseudo-color is used, the actual codes used in barcodes are not limited to colors. Any symbol or combination of symbols can be used as codes in a barcode so long as such symbol or combination of symbols are unique. Exemplary symbol or combination of symbols include but are not limited to colors, numbers, letters, shapes, or combinations thereof. It would be understood that a color based scheme would be preferred because

[00239] At step 308, once the total number of symbols to be used in a pseudo-color barcoding scheme and the number of barcoding rounds are determined, non-redundant and

unique barcodes will be created. For example, in a n-component barcode using S pseudo-color symbols, a barcode can be expressed as: $\{B_1, B_2, \ldots, B_n\}$, where each of B_1, B_2 through B_n is selected from the S pseudo-color symbols. Standard code design algorithms can be used such that the results barcodes are unique.

[00240] During hybridization analysis, barcodes can be shortened due to loss of data from one or more rounds of hybridization, For example, after the loss of one round of hybridization, n-component barcodes can become n-1- component barcodes.

[00241] When error correction algorithm is not implemented during probe design, the resulting n-1- component barcodes may now include redundant codes, making it difficult or impossible to decode targets based on the coding scheme useless. In a more specific example using colors as symbols in a pseudo-color coding scheme, two distinct six-component barcodes: {Red-Blue-Green-Red-Yellow-Blue} and {Red-Blue-Green-Green-Yellow-Blue} can become {Red-Blue-Green-~~Red~~-Yellow-Blue} and {Red-Blue-Green-~~Green~~-Yellow-Blue} if data from the fourth hybridization round becomes unavailable. As a result, the results from the entire set of experiments will be discard, wasting time and resource.

[00242] The current disclosure also creates barcodes that are drop or error-resistant. In some embodiments, one or more rounds of error correction can be implemented. Additional code design algorithms can be applied such that the resulting barcodes are error resistant; for example, the loss of any one or more rounds of hybridization data would still lead to shortened barcodes. However, the shortened barcodes are still unique and can be used to specifically identify a molecular targets.

[00243] In some embodiments, implementing an error correction mechanism involves removing barcodes that are vulnerable to loss of data. As such, the number of targets that can detected using the error-corrected scheme will be lower than that of the original scheme without error correction. To detect the same number of genes, more hybridization rounds would generally be needed.

[00244] The current method is advantageous in that it does not require additional hybridization rounds. Instead, coding capacity of the same number of hybridization rounds can be increased by designating more pseudo-color symbols.

[00245] At step 308, barcodes for an entire set of hybridization experiments will be created each corresponding to a particular molecular target. For example, a set of hybridization experiments including includes 4 barcoding rounds, each of which in turn including includes 4 serial hybridization experiments. In each hybridization experiment, 3 different detectable signals (such as three different fluorophores) will be associated with probes that bind to a set of molecular targets. Here, each detectable signal in the same serial hybridization experiment can be associated with multiple molecular targets; for example, hundreds or even more molecular targets. For example, 100 or more molecular targets can be assigned barcodes that start with the same pseudo-color symbol; i.e., all the molecular targets can be labeled with the same detectable signal. The unique identity of each molecular target will be reflected by symbols used in the subsequent three positions in the barcodes. To complete the barcoding in the example above will require that each target 3 molecular target will have a total of 48 hybridization experiments in four barcoding runs (see, Figure 10). These 48 hybridization experiments can detect a total of 20,736 molecular target (12^4) based on 12 pseudo-colors. If one error correction round is implemented, only up to 1728 (12^3) molecular targets can be detected. If 1728 different mRNA transcripts are to be detected with error correction, 1728 unique barcodes with embedded error-correct mechanism for one round of data loss will be created. Actual hybridization experiments will be carried out based on these predesigned barcodes.

[00246] At step 310, after unique barcodes are generated, each identified molecular target will be assigned one of the unique barcodes. For example, for a five component barcode based on 12 pseudo-color symbols, the barcodes will be generated using the 12 symbols. The first code can be any one of the 12 symbols, but each subsequent symbol and each subsequent barcode will be generated taking into consideration the symbols and barcodes that have already been used in order to avoid redundant barcodes. In some embodiment, additional error correction algorithms are implemented in barcode design such that the resulting barcodes are resistant to mistakes such as the loss of data from an entire barcoding round.

[00247] In some embodiments, probes can be synthesized according to the predetermined probe designs. Each probe can include a binding sequence that specifically target a site in an intended molecular target. For each molecular targets, multiple probes with different binding sequences targeting different sites of the same molecular target can be synthesized. When

probes with different binding sequences targeting the same molecular target are used in one hybridization rounds, the probes will be associated with the same detectable signal.

[00248] In some embodiments, a probe can include one or more readout sequences that can be connected, directly or indirectly, to either side of the binding sequence. In some embodiments, one or more readout sequences are directly connected to the binding sequence as one or more overhangs. In some embodiments, one or more readout sequences are indirectly connected to the binding sequence of the probe via one or more intermediate molecules; for example, as one or more overhangs of a bridge probe. Exemplary intermediate molecules include but are not limited to an RNA bridge probe, a DNA bridge probe, a protein bridge probe, a probe for hybridization chain reaction (HCR), a hairpin nucleic acid probe, an HCR initiator, an HCR polymer, other known amplification methods, or combinations thereof.

[00249] In some embodiments, the number of different readout sequences is the same as the number of barcoding rounds. In some embodiments, the number of readout sequences is greater than a number of barcoding rounds. In some and body months, each type of readout sequence, when activated, is capable of generating one color signal. In some embodiments, different readout sequence, when activated, are capable of generating the same color signal. Any detectable signals suitable for hybridization experiments can be used.

[00250] For five barcoding rounds, five groups of probes can be designed and synthesized for each molecular target. Probes in each group would increase food binding sequences that target specific sequences within a molecular target. In some embodiments, probes in the same group will bind to the same target sequence. In other and more preferred embodiments, probes in the same group will bind to multiple target sequences. As disclosed herein, each barcoding around 1 to 100 probes can be used targeting a specific molecular target. In some embodiments, 100 or fewer, 80 or fewer, 60 or fewer, 50 or fewer, 40 or fewer, 30 or fewer, 20 or fewer, or 10 or fewer probes can be used. In some embodiments, between 5 and 40 probes are used.

[00251] In some embodiments, these probes will bind to two or more target sequences, three or more target sequences, four or more target sequences, five or more target sequences, or six or more target sequences in the same molecular target. And some embodiments, the probes will be concentrated in one area of the molecular target. In some embodiments the probes will be spread out along the entire length of the molecular target.

[00252] At step 310, all probes required to carry out the entire set of hybridization experiments will be synthesized, each associated via direct or indirect connection to a primary binding sequence. For example, a set of hybridization experiments including five barcoding round each including 12 serial hybridizations will have a total of 60 hybridization experiments. After step 310, all probes required for each hybridization will be designed and synthesized.

[00253] At step 312, serial hybridization experiments can be performed using probes synthesized in the previous step. Barcodes associated with these probes are stored in a database or library. Each probe will be associated with a code in the designated barcode for the molecular target of the probe. The number of hybridization experiments within a serial hybridization round will be determined based on the availability of different types of detectable signals (such as colors) and the total number of symbols in the pseudo-color barcoding scheme. As an example, hybridization experiments based on a 60 pseudo-color scheme will be carried out s using 4 different types of color signals. Then 5 rounds of barcoding with the 60 pseudo-color will be carried out for $60^5 = 7.776 \times 10^8$ number of barcodes. In some embodiments, in each barcoding round, 15 serial hybridization rounds will be performed. In each of the 15 hybridization round, all four color signals are used to each identify a different molecular target. Here, after the particular round of serial hybridization experiments, 60 unique molecular targets will be labeled with color signals. In some embodiments, 20 serials hybridization arounds will be performed using three of the four color signals, each color representing a different molecular target. Again, after the particular round of serial hybridization experiments, each 60 unique molecular targets are decoded and associated with a previously assigned pseudo-color symbol. In some embodiments, any combinations of three colors can be used in the serial hybridization experiments. In some embodiments, in each hybridization round, two to four color signals can be used, each representing each representing a different molecular target. In some embodiments, one color signals can be used in one or more hybridization experiments in a serial hybridization round. Here, one color must be present in separate hybridization experiments in order to represent different molecular targets. For example, if one is limited only one type of color signal, hybridization experiment can be carried out 60 times in each barcoding round. Each hybridization experiment identifies only one molecular target. It would be understood that performing a high number of hybridization experiments may not be desired because the

biological sample being analyzed can be degraded, for example, by enzymes that digest nucleic acids.

[00254] In some embodiments, during each hybridization experiment in each serial hybridization round, reference visual signals are introduced. These reference visual signals are associated with one or more alignment references that are associated with a biological sample and do not change their positions between hybridization rounds. These alignment references can serve as standard for subsequent image analysis. They can be part of the biological sample or external material added to and immobilized with the biological sample. In some embodiments, alignment references are immobilized at the same time as the biological sample. In some embodiments, alignment references are immobilized at a different time from when the biological sample is immobilized. Exemplary alignment references include but are not limited to beads, oligonucleotide sequences immobilized on the coverslips and detected by a complementary oligo, microscopic objects (e.g., a metal bead, a gold bead, a polystyrene bead loaded with a dye), PCR handle sequences on the primary probe, or combinations thereof. As disclosed herein, the reference signals are fiducial markers for alignment between the different serial and barcoding hybridizations. They are present throughout all of the hybridizations. In some embodiments, at least one marker is present per image. They can be probes targeting all of the primary probes (as shown in Figure 18a, yellow probes), such that all the dots are observed in all of the hybridizations. In some embodiments, the reference signals can be produced from beads attached to the cover glass, or other molecules or particles. In some applications, where "super-resolution" capabilities are used, signals form the beads can be Gaussian fitted to nanometer resolution, and images between hybridizations can be aligned to that precision. In some embodiments, this becomes useful for the in situ application where Gaussian fitting allow more molecular targets to be detected in each pseudo-color and then collapsed into the composite image for barcoding.

[00255] At step 314, an image of visible probes bound to the biological sample will be taken after each hybridization experiment in a serial hybridization round. All images from hybridization experiment in a serial hybridization round will be aligned using reference visual signals to create a composite image.

[00256] At step 316, it will be determined whether there are one or more barcoding rounds remaining based on the barcodes pre-designed at step 308 If yes, the method returns to steps 312 and 314 and more serial hybridization experiments are carried out to constitute another barcoding round and produce another composite image.

[00257] If no, the hybridization portion of the analysis is complete and the method proceeds to step 318. Images and composite images will be subject to further data analysis. For example, in a sample with limited space, e.g., in a single cell, Gaussian localization analysis can be carried to generate, for example, more focused color spots each representing binding to a molecular target. Other data processing methods can also be used to enhance data quality, for example, a de-noise mechanism can be applied to reduce back signals.

[00258] The method concludes at step 320.

[00259] Figure 4 depicts a schematic overview of Sequential Probing Of Targets (SPOTs), targeting RNA samples using a pseudo-color barcode scheme. The method performs transcriptome level profiling of mRNAs with single molecule sensitivity and high accuracy using a method based on sequential FISH (seqFISH) [Lubeck 2014]. The initial demonstration of seqFISH showed promising coding capacity to barcode a large number of molecular species in cells and tissues [Shah 2016]; however, a major limitation has been identified: when seqFISH is performed in cells, optical diffraction limit prevents many mRNAs to be resolved simultaneously due to the limited space in a cell.

[00260] As shown in Figure 4, there are a total of 5 barcoding runs and each barcoding run has four serial hybridizations. Three different fluorophores (e.g., red, green and blue) are used in each serial hybridization experiment. For simplicity, each serial hybridization experiment in Figure 4b only shows one molecular target for each fluorophore. However, in practice, each fluorophore can represent tens, hundreds or even more molecular targets. In addition, unlike the illustration in Figure 4b which shows that the same number of targets are tagged in each serial hybridization, different numbers of molecular targets can be tagged in different serial hybridizations. For example, for a 1,500 mRNA experiment of 4 barcoding runs, serial hybridization No. 1 of barcoding round 1 can tag 400 mRNA transcripts with three different fluorophores (e.g., red, green and blue). Serial hybridization No. 2 of the same barcoding round can tag 280 mRNA transcripts with the same three fluorophores. 320 mRNA transcripts can be

tagged the same three fluorophores in serial hybridization No. 3. And 500 mRNA transcripts can be tagged the same three fluorophores in serial hybridization No. 4. As the numbers suggest, within each serial hybridization, multiple mRNA transcripts can be tagged with the same fluorophore. In some embodiments, a fluorophore is attached to a probe via multiple readout sequences, which in turn are connected to a primary binding sequence targeting a specific site in a molecular target. In some embodiments, the readout sequences can be connected directly to the primary sequence. In some embodiments, the readout sequences can be connected indirectly to the primary sequence, for example, via one or more intermediate molecules. In some embodiments, the one or more intermediate molecules comprise an RNA bridge probe, a DNA bridge probe, a protein bridge probe, a probe for hybridization chain reaction (HCR), a hairpin nucleic acid probe, an HCR initiator, an HCR polymer, or combinations thereof.

[00261] In addition, there can be different number of serial hybridizations in two different barcoding rounds. For example, in the example presented above, 4 serial hybridizations are included in barcoding round 1. Barcoding round 2 can have 4 or 6 or 3 serial hybridizations, so long as the results from the two sets of serial hybridizations can reveal the same 1,500 mRNA transcripts. For example, images from all serial hybridizations in the same barcoding round can be compiled into a composite image for the barcoding round. In this barcode scheme, 12 pseudo-colors are created, which can be assigned numbers as 1 through 12. In some embodiments, 12 different colors are assigned as symbols of the 12 pseudo-colors. After all 5 barcoding rounds are completed, the 1,500 mRNA transcripts will be fully coded.

[00262] In some embodiments, the total number of targets revealed in two different barcoding rounds can differ. For example, there are some problems and only 1,400 mRNA transcripts are tagged in in barcoding round 1 while 1,350 mRNA are tagged in barcoding run 2. Subsequent image and data analysis can reveal that only 1,200 of the mRNA transcripts overlap between the two barcoding rounds. And ultimately only 1,200 mRNA transcripts are fully coded in all 5 barcoding rounds.

[00263] The pseudo-color scheme can overcome this density problem in both in situ and in vitro applications. In the in vitro cases, implemented with SPOTS, capturing transcripts onto an oligonucleotide dT surface and adjusting the dilution factors can easily remove the optical crowding problems and allow the transcriptome to be imaged by seqFISH.

[00264] In addition, pseudo-color is more efficient in terms of imaging time than all existing imaging methods including expansion microscopy. For example, with 3 fluorophores, it takes $3^{(10-1)}$=19,683 to code for the transcriptome with one round of error correction. Thus, a total of 30 frames of imaging is required. In a pseudo-color scheme, 20 pseudo-colors can be used for 4 rounds of hybridization to code for $20^{(4-1)}$=8000 genes in each of the three fluorophore channels for a total of 24,000 genes. This requires 3x20x4=240 frames to image, a 8 fold increase in the imaging time. However, because the density of mRNA is effectively diluted into 3x20=60 pseudo-channels instead of 3 fluorophore without pseudo-color coding, the density problem is alleviated with a factor of 20. The target spots can be localized to nanometer precision by Gaussian fitting to decrease the density in each pseudo-color channel before barcode alignment. Thus, the benefit of pseudo-color coding is to decrease the density of the target spots in the cell, while saving imaging time compared to expansion microscopy, where expanding the sample by 20 folds requires an additional 20 fold increase in imaging time. Since imaging time is rate limiting in any sequential imaging method, pseudo-color solves a major problem in implementing transcriptome profiling in situ.

[00265] In some embodiments, for in situ experiments based on a pseudo-color scheme, fewer number of different detectable signals can be used so that they can be resolved from each other more clearly due to the limited amount of real estate within a cell. In some embodiments, with super resolution microscopy, even densely populated detectable signals can be resolved from each other.

[00266] As an example, three different color signals are used in an in situ experiment. In some embodiments, for each color, hybridization experiments can be repeated 20 times to generate a 20 pseudo-color image. Here, each 20 pseudo-color image is compiled from only 1 real color channel although multiple colors can be used. The single-color approach offers several advantages during imaging analysis. Significantly, there are chromatic aberrations in a microscope, and aligning images across different actual color channel can be difficult and lead to errors or inaccuracies. In the current embodiments, all 20 pseudo-color can be from the same color channel (e.g., the Cy5 channel) then all images are by definition aberration free.

[00267] Using an mRNA hybridization experiment as an example, mRNA transcripts bound to activated color probes are visualized as dots in the image. In some embodiments,

Gaussian fitting is applied to localize the RNA dots in cells and generate a higher resolution image or less dense image of the cell. In some embodiments, Gaussian localization can take place at one or more time points; for example, for each image corresponding to a serial hybridization experiment, for the images from all serial hybridization experiments in a barcoding round; for each composite image corresponding to a barcoding round. In some embodiments, Gaussian localization can be repeated multiple times.

[00268] In some embodiments, color beads can be used to align the pseudo-color images to create composite images.

[00269] When multiple color probes are used (e.g., 4 or 5 colors), the resulting image can be too dense and it can be difficult or impossible to apply the Gaussian fitting process, which requires discrete dots. By virtually diluting the samples to 60 pseudo-color, then the density becomes much less of an issue. After the first barcoding round, DNAse can be used to strip the probes, which carries the RNA targeting sequence and one round of readout sequence. This ensures that when barcoding is done in cells (in situ), there is no nonspecific binding. A new set of probes with the 2nd barcode sequence is then hybridized. Another 20 rounds of serial hybridization with 3 colors is done. In some embodiments, images from the hybridization experiments are subject to Gaussian fit, and be compiled into 3 sets of 20 color pseudo-color images.

[00270] This process can be repeated 4 times to generate a 20x20x20x20 barcode in each color. So 8000 error corrected barcodes are available in each color, to encode a total of 24000 barcodes.

[00271] In this example, the barcode scheme can be actually 3 separate pieces of 8000 barcodes in each of the colors. There can be 60 readout sequences total. Because of the DNAse stripping step, the readout sequences used in the 2nd barcoding hybridization can be the same as 1st barcoding hybridization.

[00272] For in vitro analysis, all the readout sequences can be added onto the primary binding probes. In those cases, the readout sequences for the different barcoding hybridizations have to be different. The approach is possible in vitro because there is very little nonspecific binding, since the biological sample has often been extracted out of the cells. For in situ, more stringent conditions are needed to avoid nonspecific binding in cells.

[00273] An alternative approach is to generate 100 readout sequences all in one color and just code for 10,000 targets with 3 rounds of hybridization (100x100x100) with one round of error correction. As disclosed herein, any manageable number of readout sequences can be used; for example, 50 or fewer, 60 or fewer, 80 or fewer, 100 or fewer, 150 or fewer, 200 or fewer, 250 or fewer, 300 or fewer, 400 or fewer, or 500 or fewer. In some embodiments, more than 500 readout sequences can be used.

[00274] As disclosed herein, the pseudo-color scheme is better than expansion microscopy in solving the density issue. Expansion microscopy is a new technology for imaging biological samples with fine detail by physically making the samples bigger through a chemical process that preserves nanoscale isotropy. Expansion microscopy enables super-resolution imaging on a conventional light microscope. However, expansion is a complicated procedure and the sample can contract and expand during the rehybridization process, which makes aligning the barcodes hard. Also, if the tissue sample is expanded 10 times, the tissue will be imaged 10 times longer. So even though pseudo-color appears to require more hybridizations, the total imaging time is comparable or even more favorable compared to if one had to expand the sample. For example, the 60 pseudo-color scheme is basically 20x expansion compared to just a 3 color scheme.

[00275] In a seqFISH experiment, 10 rounds of hybridization are needed to encode 20,000 genes with error correction with 3 colors ($3^{(10-1)}$=19,683). As disclosed herein, the pseudo-color scheme use 20x4=80 rounds of hybridizations for each channel. Thus, the number of hybridizations gone up by 8 fold, but the density is diluted down by a factor of 20. Thus the current approach is more efficient than expansion microscopy. For example, one can also use 1 channel and 100 pseudo-color and three rounds of hybridization, which correspond to 33.3 fold (100 pseudo-colors/3 colors) improvement in dilution and only 10 fold increase in imaging time. As disclosed herein, 3 color signals and 20 pseudo-color in each channel have produced excellent results. In practice, one can balance the "dilution" factor with the number of targets being analyze to determine the number of the actual color signals and the number of pseudo-colors.

[00276] In some embodiments, the current approach can also be combined with correlation FISH (corrFISH, Coskun and Cai, Nature Methods 2016) to quantify the abundances of highly expressed mRNAs. corrFISH can be applied to decoding samples where the number of

target dots are so high that they are individually resolvable. This approach can decode the abundance of the genes, but trade off in spatial resolution. As an example of implementing the corrFISH scheme in addition to the existing pseudocolor scheme, a 4th channel can be dedicated to barcode highly abundant transcripts by corrFISH in addition to the 3 colors channels used for generating the pseudo-colors. The 4th channels can generate 20 pseudo-color through 20 rounds of serial hybridization. Then corrFISH can be used to code for another 8000 genes, but highly expressed.

[00277] In some implementations, primary probes can be stripped and signals extinguished after the pseudocolor schemes and probes targeting another set of targets can be hybridized. The utility of this can be that multiple types of targets can be probed sequentially in the same cells. For example, 20,000 intronic RNAs can be first targeted through the pseudo-color scheme. Then 20,000 mRNAs can follow, with 20,000 proteins after that. Or alternatively, probes can be used to target specific highly expressed genes to measure their localizations in cells, especially cells with polarization, such as processes in neurons. Or probes can be used to target specific isoforms or combination of isoforms of specific targets.

[00278] As disclosed herein, the approached disclosed herein can also work with amplification methods, such as HCR, to implement this work robustly in tissues.

[00279] It will be noted that although the approach above is described in connection with an in situ experiment, one of more aspects of it can be applied to in vitro experiments as well.

[00280] In one aspect, disclosed herein are readout probes with cleavable linkers. Figure 25 depicts exemplary chemical reactions for synthesizing a readout probe with a disulfide linker.

[00281] In one aspect, sequential barcoding FISH (seqFISH) is performed by using nucleic acid readout probes that are conjugated with a signal moiety via a cleavable linker. Any suitable cleavable linkers can be used, including but not limited to an enzyme cleavable linker, a nucleophile/base sensitive linker, reduction sensitive linker, a photo-cleavable linker, an electrophile/acid sensitive linker, a metal-assisted cleavable linker, or an oxidation sensitive linker. Exemplary linkers can be found in Leriche et al., 2012, "Cleavable linkers in chemical biology," Bioorganic & Medicinal Chemistry 20:571–582, which is hereby incorporated herein in its entirety.

[00282] In some embodiments, the cleavable linker is a disulfide linkage. In some embodiments, the cleavable linker is a nucleic acid restriction site. In some embodiments, the cleavable linker is a protease cleavage site.

[00283] An exemplary system utilizing nucleic acid readout probes is shown in Figure 26. As depicted, a gene specific primary probe binds to a target site; e.g., in an mRNA molecule under an in situ or in vitro setting. Besides a binding sequence, the primary probe further includes an overhang sequence at one end of the binding sequence. In some embodiments, a second overhang sequence is included at the other end of the binding sequence.

[00284] In some embodiments, an overhang sequence includes one or more target sequences to which one or more nucleic acid readout probes bind. In some embodiments, each target sequence uniquely interacts with a set of readout probes with specific readout binding sequences. As disclosed herein, an overhang sequence may include two target sequences, three target sequences, five or fewer target sequences, seven or fewer target sequences, or ten or fewer target sequences. In some embodiments, an overhang sequence may include ten or more target sequences. Similar arrangements can be implemented where there are two overhang sequences.

[00285] In some embodiments, an overhang sequence binds to a bridge probe that provides target sequences for one or more readout probes to bind, as depicted in Figure 26. A bridge probe can be interchangeably called an intermediate bridge probe or a secondary bridge probe. A bridge probe includes a binding sequence that binds to all or a portion of an overhang sequence in a primary probe. In some embodiments, a bridge probe further includes one or more readout binding targets that are connection in series and linked to the binding sequence.

[00286] In some embodiments, as depicted in Figure 26, two bridge probes can bind to the same primary probe via two overhang sequences.

[00287] As disclosed herein, a bridge probe may include two readout binding targets, three readout binding targets, five or fewer readout binding targets, seven or fewer readout binding targets, or ten or fewer readout binding targets. In some embodiments, an overhang sequence may include ten or more readout binding targets. Similar arrangements can be implemented where there are two bridge probes bound to overhang sequences.

[00288] Exemplary rehybridization schemes utilizing the readout probes are illustrated in Figure 26. For example, the first round of rehybridization (hyb1) begins with the hybridization

of gene specific primary probes to the target mRNA. Each gene specific primary probes contains one or more "overhang" sequences that allows the secondary bridge probes to hybridize against. The secondary bridges contain two or more tertiary readouts binding sites which is the key to efficient and quick rehybridization. In the first hybridization, unique tertiary readout probes conjugated with blue dye are hybridized to their unique binding sites on the secondary bridge probe. Once imaged, the sample is treated with reducing agent such as TCEP or DTT to cleave off the disulfide-linked dyes. Then, the sample is washed with wash buffers. During the second round of hybridization, a second set of unique tertiary readout probes with red dye is hybridized to its unique binding site on the secondary bridge. After two rounds of hybridizations, a particular mRNA is then barcoded with a color barcode of red and blue. Additional rounds of hybridization can be applied to create more sophisticated barcoding sequences. Technically, the scaling factor of seqFISH with this rehybridization method depends on the number of available secondary bridges with its number of unique tertiary probes binding sites. For example, by incorporating 2 secondary bridges with total 8 unique tertiary readout binding sites (N=8) , and with 4 fluorophores (F=4), one can generate up over 64,000 unique barcodes ($F^N = 4^8 = 65,536$). Moreover, in embodiments where bridge probes are used, it is possible to strip off the secondary bridges with high concentration of formamide, and flow in another unique set of secondary bridges to continue the scaling process, which further increases the upper limit of the scaling factor.

[00289] In one aspect, disclosed herein are methods and systems for amplifying visual signals during each round of hybridization during sequential hybridization reactions, based on hybridization chain reaction (HCR). An exemplary embodiment of HCR is illustrated in Figure 27A. During hybridization round 1, probes with overhang initiator sequences are added to a nucleic acid target molecule such as an mRNA or a DNA. Also added are hairpin nucleic acid probes bearing sequences complementary to those of the initiator sequences. The presence of initiator sequences cause unfolding of the hairpin nucleic acid probes and result in chain reactions that lead to self-assembled extended HCR polymers. Because each hairpin nucleic acid probe bears a signal, self-assembled extended HCR polymers result in amplification of signals and better detection of target sites.

[00290] Figure 27B illustrates an exemplary readout probe embedded with a cleavable linker. Here, the cleavable linker is a disulfide bond. At one end of the cleavable linker, a readout probe as disclosed herein includes a binding sequence that allows it to bind to a specific nucleic acid target. In some embodiments, the nucleic acid target is an mRNA or a DNA. In some embodiments, the nucleic acid target is within an intact cell or as part of cell extract. In some embodiments, the nucleic acid target is within a primary binding probe that directly binds to a target site in an mRNA. In some embodiments, the nucleic acid target is within a secondary binding probe that binds to a primary binding probe that directly binds to a target site in an mRNA. In some embodiments, the nucleic acid target is within a tertiary or quaternary binding probe. One of skill in the art can apply the principle to any level of binding and interaction.

[00291] At the other end of the cleavable linker, a readout probe as disclosed herein further includes an HCR initiator sequence. When exposed to hairpin nucleic acids bearing partial or complete complementary sequences, the initiator sequence can trigger a chair reaction that allows a signal motif formed by multiple extender probes. Each extender probe includes a signal moiety. Aggregation of multiple extender probes enhances signal detection.

[00292] An exemplary scheme for forming a signal motif with multiple extender probes during a sequential hybridization process is illustrated in Figure 27C. During the first round of hybridization, nucleic acid detection probes with embedded cleavable linkers binds to a first target site within a nucleic acid target sequence. In some embodiments, extender probes are added after the initial binding of nucleic acid detection probes to the first target sequences. In some embodiments, extender probes form an aggregate before the aggregated polymer is added to the reaction mix and binds to the imitator sequence in the nucleic acid detection probes.

[00293] In some embodiments, extender probes are standard hairpin probes each including a sequence that is partly or completely complementary to the initiator sequence in the readout probes. In these embodiments, extender probes are very similar or identical to each other. The size of the resulting extendible signal motif may be controlled by the concentration or absolute quantity of the extender probes added.

[00294] In some embodiments, extender probes including different types of nucleic acid sequences can be used to achieve controlled signal amplification. For example, the signal can be amplified five times if five populations of extender probes are used: {EP_1, EP_2, EP_3, EP_4, and

EP_5}. The first population of extender probes includes a binding sequence that binds to all or a part of the initiator sequence. The second population of extender probes includes a binding sequence that binds to a region in the first population of extender sequence. The third population of extender probes includes a binding sequence that binds to a region in the second population of extender sequence. The fourth population of extender probes includes a binding sequence that binds to a region in the third population of extender sequence. The fifth population of extender probes includes a binding sequence that binds to a region in the fourth population of extender sequence. In such embodiments of linear amplification, the size of the resulting extendible signal motif can be controlled by the number of populations of extender probes that are provided.

[00295] In some embodiments, an extender probe may include multiple binding sites for binding subsequent extender probes. For example, besides binding to the initiator sequence, EP_1 may include two or more binding sites for EP_2, thus allowing further amplification of the signal. This form of amplification may occur at any level. For example, in the example above, multiple binding sites for subsequent or downstream extender probes can be implemented in any one or combinations of EP_1, EP_2, EP_3, or EP_4. For example, extender probes from EP_2, EP_3, or EP_4 can all bind to target sites in EP_1, which in turn binds to the initiator sequence.

[00296] In some embodiments, the amplification occurs at multiple levels. Generally, when m populations of extender probes are present, multiple binding sites for subsequent or downstream extender probes cam be implemented in any one or combinations of EP_1, EP_2, …, or EP_{m-1}. Additionally, when multiple binding sites are present, they can be connected in series or arranged in a non-linear fashion (e.g., in a branched or circular arrangement). Depending on the number and configuration of the binding sites, the resulting extendible signal motif can be a stick, a ball, a net or in any other applicable form.

[00297] One of skill in the art would understand that any suitable number of populations of extender probes can be added to achieve an optimal signal to noise ratio for the best imaging effects. For example, the extender probes can include five or fewer, seven or few, 10 or fewer, 15 or fewer, 20 or fewer, 25 or fewer, 30 or fewer, 40 or fewer, 50 or fewer populations.

[00298] In some embodiments, the extender probes are mixed together prior to being mixed with the readout probes having the initiator sequence. In some embodiments, the extender

probes are sequentially added to the readout probes having the initiator sequence where the readout probes are already bound to its nucleic acid targets.

[00299] As shown in Figure 27C, after imaging analysis, a cleaving agent can be applied to sever the linker between the binding sequence and the imitator sequence in a readout probe. The amplified polymers can then be cleaved off and washed away.

[00300] During a second round of rehybridization, new nucleic acid detection probes are applied. The new nucleic acid detection probes include a different binding sequence that binds to a second and different target site in the nucleic acid target sequence. The new nucleic acid detection probes also include a cleavable linker and an initiator sequence. The initiator sequence can be the same as or different from the initiator sequence from the previous set of nucleic acid detection probes.

[00301] The new extender probes are used, as described hereinabove, to form amplified polymers to enhance signal detection. After imaging analysis, the new set of amplified polymers can be cleaved off and washed away. By using extender probes bearing a different type of visual signals, barcodes can be established for nucleic acid targets. Depending on the availability of target sites within a nucleic acid target, multiple rounds of hybridizations can be performed to create more complex barcodes. For example, there can be three rounds of hybridizations, four rounds of hybridizations, five rounds of hybridizations, seven or fewer rounds of hybridizations, 10 or fewer rounds of hybridizations, 12 or fewer rounds of hybridizations, 15 or fewer rounds of hybridizations, 20 or fewer rounds of hybridizations, 30 or fewer rounds of hybridizations, 40 or fewer rounds of hybridizations, or 50 or fewer rounds of hybridizations.

[00302] The compositions and methods disclosed herein can be used in sequential hybridizations to identify any suitable cellular targets within an intact cell or in an *in vitro* setting. In some embodiments, the cellular targets can be mRNAs or DNAs. In some embodiments, the cellular targets can be proteins. For example, the initial target-binding primary probe can be an antibody conjugated with nucleic acid sequence for subsequent bindings.

[00303] As exemplified herein, provided technologies work for a wide variety of samples. For example, HCR-seqFISH worked in brain slices and that SPIMs can robustly detect single mRNAs in CLARITY brain slices. In some embodiments, provided technologies are useful for

profiling targets in mouse models of neurodegenerative diseases, or human brains. No other technology prior to the present invention can deliver the same quality and quantity of data.

Error Correction Mechanism

[00304] Figure 3A illustrates general aspects of a sequential hybridization analysis that may contribute to quality of the analysis. Sequential hybridization includes multiple rounds of hybridization, where each round of hybridization is a multiple step process. Errors can be introduced at any step during any round of hybridization. Such errors can lead to misidentification of target genes in a sample.

Barcodes and Error Correction

[00305] In one aspect, disclosed herein are methods for designing barcodes with built-in error correction mechanisms such that the multi-component barcodes can withstand the loss of the data from one or more rounds of hybridization (i.e., drop-safe). As disclosed herein the terms "barcode" and "code" are used interchangeably.

[00306] As disclosed herein, by using probes that are associated with F detectable visual signals ($F \geq 2$), a sequential hybridization of N rounds ($N \geq 2$) can generate a total of F^N combinations of visual signals. In some embodiments, these combinations of visual signals can be used as barcodes to uniquely identify cellular targets such as mRNA, DNA, or even protein.

[00307] Figure 30 illustrates an exemplary process 3000 for generating drop safe barcodes.

[00308] At step 3010, the total number of genes that will be analyzed during the hybridization experiments is determined. This number sets the threshold values for the number of detectable visual signals (F) and the total number of rounds in the sequential hybridization (N).

[00309] Once the total number of genes is determined, steps 3020 and 3030 are performed simultaneously. The number of genes being analyzed must be smaller than the total number of possible combinations of visual signals (F^N). Practical aspects of the hybridization analysis need to be considered when selecting values for F and N. One would tend to reduce the number of rounds of hybridization to as few as possible. Theoretically, this can be achieved by using a high number of detectable visual signals (F). In practice, however, too many different types of visual

signals may interfere with each other. For example, overlapping of visual signals can lead to barcode misidentification.

[00310] At step 3040, a library of drop-safe unique barcodes are generated by implementing one or more error correction mechanisms.

[00311] In some embodiments, a repeat round can be performed for any round during a sequential hybridization of N rounds, rendering a new sequential hybridization of $(N+1)$ rounds. The extra repeat round can be an error correction round. The repeat round can be a duplicate of any round of the n rounds sequential hybridization. The repeat round can take place as any round during the sequential hybridization $(N+1)$ rounds.

[00312] After the repeat, there are two rounds of hybridization that should be identical to each other. Consequently, the complete loss of one of the repeat rounds does not affect the outcome of the sequential hybridization. As such, either of the repeat rounds is a drop-safe round.

[00313] Figure 2 illustrates an experiment where 3 rounds of hybridization using probes with 4 types of detectable visual signals (red: R, yellow: Y, green: G, and cyan: C) are used to create barcodes for 4 different mRNA molecules. Hybridization round 3 is a repeat of hybridization round 1, as summarized in Table 1 below.

Table 1. Illustration of the effect of repeat hybridization rounds.

mRNA molecules	Color barcodes (3 rounds)	Color barcodes (dropping round 1)	Color barcodes (dropping round 3)
mRNA1	Y-C-Y	C-Y	Y-C
mRNA2	G-R-G	R-G	G-R
mRNA3	R-C-R	C-R	R-C
mRNA4	C-R-C	R-C	C-R

[00314] As shown in the table above, data from one of the repeat rounds can be dropped completely in case of major experiment error, barcodes derived from the remaining rounds of hybridization still uniquely represent the mRNA molecules.

[00315] In some embodiments, even in a questionable hybridization round, most of the information is still reliable. Only some of the bindings between probes and target sequences include inaccurate information. In some embodiments, partial data from a questionable round of hybridization are used. For example, in the illustration above, binding signals can be missing or ambiguous for a particular location during hybridization round 1, which can produce an incomplete three letter barcode *-C-Y for the particular location, where * remains undetermined. In the scheme illustrated, the identity of * is not needed to decipher that the code is for mRNA1. Similarly, binding signals can be missing or ambiguous for a particular location during hybridization round 2, which can produce an incomplete three letter barcode R-*-R for the particular location, where * remains undetermined. Once again, the identity of * is not needed to decipher that the code is for mRNA3.

[00316] Additionally, data from repeat rounds can validate each other. For example, in Figure 2C, a circle highlights a cyan data point in the image corresponding to hybridization round 2. In the same location, the image corresponding to hybridization round 3 reveals a yellow data point. Based on only information from hybridization rounds 2 and 3, this location would be identified as part of mRNA1. However, no signals are identified at the location during hybridization round 1, which suggests that the highlighted data points may be due to non-specific binding.

[00317] In some embodiments, a sophisticated barcode generating algorithm is used such that the resulting barcodes can withstand the loss of any round or even multiple rounds of hybridization data. In some embodiments, a barcode generator is used to generate the drop-safe barcodes. For example, Figure 32 illustrates an example, where probes with 5 different visual signals (blue, green, red, purple and yellow) are used in 4 rounds of hybridization. One of the hybridization round is an error correction round where barcodes are generated based on barcodes from the previous 3 rounds. The following is an example that illustrates how barcodes are generated.

[00318] Designing an error correction code to correct for m number of errors in a message of n length is analogous to packing as many spheres of radius m in a n dimensional cube. There are examples of "perfect codes" such as Golay and Hamming codes that can be as efficient as possible in this packing design. These perfect codes are important in digital communication

because the word lengths are long, up to billions of letters for gigabytes of data, and many forms of errors can occur, including deletion and insertions. However, in the seqFISH experiments, as the code lengths are short, a perfect code correction system is not necessary, especially as the "correct" codes are already defined. One of the major source of error is deletions due to loss of a hybridization. Thus, it is possible to design simple correction schemes that are not completely efficient (i.e. obtain the tightest packing density for the n-spheres) but can achieve good error correction with just a few extra rounds of hybridization.

[00319] To design a barcode scheme that can tolerate loss of a single round of hybridization is akin to a problem where any n-dimensional hypercube is collapsed by 1 dimension to a n-1 dimensional hypercube without having any two points on the n-dimensional hypercube mapping to the same point. In order for this to be true, no two barcodes can be connected by a 1D line running parallel to any of the axes. There are many solutions to generate this 1 round loss tolerant code.

[00320] In this example, 4 rounds of hybridization is used. Here, 5 different visual signals (blue, green, red, purple and yellow) are assigned numerical values. In some embodiments, the numerical values are integers. For example, blue = 1; green = 2; red = 3; purple = 4; and yellow = 5. It would be understood that these are mere sample values. Any non-redundant numerical values can be assigned to represent the different types of visual signals. In some embodiments, a barcode generator is used to generate the barcodes used in the experiment. In the exemplary embodiment, a drop-safe barcode for a particular target gene can be defined as a four-component linear array: $\{i, (i + j + k) mod\ 5, j, k\}$. Here, mod (modulo operation or modulus) finds the remainder after division of one number by another. For example, 8 mod 5 is 3. 5 mod 5 is 0, which is equivalent to 5.

[00321] In this example, i represents the numerical values corresponding to the visual signals observed for the particular target gene during the first round of hybridization. The scheme $(i + j + k) mod\ 5$ represents the numerical values corresponding to the visual signals observed for the particular target gene during the second round of hybridization. j represents the numerical values corresponding to the visual signals observed for the particular target gene during the third round of hybridization. k represents the numerical values corresponding to the visual signals observed for the particular target gene during the found round of hybridization. In

this example, i, j, and k each can be 1, 2, 3, 4 or 5,or any one of the numerical values that have been assigned to the five types of visual signals used in the experiment.

[00322] In this example, $(i + j + k) mod\ 5$ is determined as the error correction round. However, once complete barcodes are generated, any of round 1 through round 4 can be dropped to yield unique 3-component barcodes. As such, the barcodes determined by this method can be used to correct errors in any round.

[00323] The following table illustrates how the 1 drop tolerant barcodes can be generated using the equation $(i + j + k) mod\ 5$.

Table 2. Illustration of the effect of repeat hybridization rounds.

Genes	**1st round of hyb***	**2nd round of hyb**	**3rd round of hyb**	**4th round of hyb $(i + j + k) mod\ 5$**
mRNA1	1	2	4	2
mRNA2	3	3	1	2
mRNA 3	5	1	2	3
mRNA 4	2	3	5	5
…	…	…	…	…
mRNA125	5	2	1	3

*The term "hyb." stands for hybridization. Numerical values are assigned to color signals as follows: blue = 1; green = 2; red = 3; purple = 4; and yellow = 5.

[00324] As illustrated above, although the 4th round of hybridization is generated using an error correction algorithm, any one round of four rounds of hybridization in Table 2 can be dropped and still yield a unique set of barcodes for 125 genes.

[00325] More generally, a barcode that can resist the elimination of one round of hybridization can be defined as:

$$\{j_1, j_2, \ldots (a_1 * j_1 + a_2 * j_2 + \cdots + a_n * j_n + C)\ mod\ F, \ldots, j_n\} \qquad (1)$$

where j_1 is a numerical value that corresponds the detectable visual signals used in the first round of hybridization, j_2 is a numerical value that corresponds the detectable visual signals used in the second round of hybridization, and j_n is a numerical value that corresponds the detectable visual signals used in the *n*th round of hybridization. In some embodiments, $j_1, j_2, \ldots j_n$ are non-redundant integers. In some embodiments, $a_1, a_2, \ldots a_n$ can be any integers that are not none zero. In some embodiments, C is a constant integer. In some embodiments, C is zero. The remainder of F divided by F is 0 (F mod F = 0), so F and 0 are equivalent. There is no limitation on the number of hybridization. One of such examples is shown in Figure 37.

[00326] Array (1) is a general representation of a barcode that is safe against the drop or loss of one round of hybridization. Although $(a_1 * j_1 + a_2 * j_2 + \cdots + a_n * j_n + C)\ mod\ F$ is the designated error correction round, in some embodiments, the barcode is safe against the loss or drop of any round of hybridization.

[00327] As disclosed herein, array (1) consists of n-component, each corresponding to the visual signals from a particular round of hybridization. In some embodiments, probes binding to a particular gene are all associated with the same detectable visual signal, for example, red, green or blue. In some embodiments, probes binding to a particular gene are all associated with multiple types of detectable visual signal, for example, green + yellow or blue + red. Through combinations of visual signals, the total number of different types of detectable visual signals can be further expanded.

[00328] In some embodiments, barcodes can be designed such that drop or loss of data from two rounds of hybridization can be tolerated. Using 2 additional rounds of hybridization does not correct for all possible 2 drops, but it does correct for a large fraction of the 2 drops. For example, for detecting 100 genes with F=5 dyes, 3 rounds of hybridization are needed for basic barcoding of these genes. When adding two rounds of hybridization, the error correction code:

$$\{i, j, k, (i + j + k)\ mod\ F, (i - j)\ mod\ F\} \quad (2)$$

[00329] Such codes can correct for 2 drops all except dropping hybridization round 3 and round 4 together. Here, each component in the 5-member array represents one round of hybridization.

[00330] Similarly, an error correction code such as

$$\{i, j, k, (i + j + k)\ mod\ F, (i - k)\ mod\ F\} \quad (3)$$

can correct for dropping hybridization round 2 and hybridization round 4 together. Again, each component in the 5-member array represents one round of hybridization.

[00331] For example, to code for most of the transcriptome, only 6 rounds of hybridization are needed when F=5 (6^5=15,625). When adding two rounds of hybridization, the following error correction code is generated:

$$\{i, j, k, l, m, n, (i + k + l + m + n)\ mod\ F, (i - j - k - l + n)\ mod\ F\} \quad (4)$$

[00332] There are a total of 28 combinations of how 2 rounds of hybridization can be lost or dropped. This type of code can correct for 24 out of the total 28 combinations. Here, each component in the 8-member array represents one round of hybridization. Similarly, the 1st error correction round can be any liner combination of 5 out of 6 rounds of hybridization (e.g., without j) and 2nd error correction can be a subset of the linear combination of 5 out of 6 rounds of hybridization (e.g., without m). In these embodiments, in the 2nd error correction round, indices include different coefficients as long as the it is not exactly the same 5 indices used in the 1st error correction round.

[00333] To correct for all combinations of drop or loss of 2 rounds of hybridization (2 drops) fully, 3 additional hybridizations are needed. Again for 6 rounds of hybridization with 5 types of detectable signals (F=5), three extra rounds of hybridizations are added to create the full 9-member error correction code:

$$\{i, j, k, l, m, n, (i + j + k + l + m + n)\ mod\ F, (i - j - k - l)\ mod\ F, (m - n - j + k)\ mod\ F\} \quad (5)$$

[00334] In some embodiments, there are many equivalent codes that can correct for 2 drops with 3 additional rounds of hybridization. They can be all empirically determined. The number of hybridization for any reasonable number can be simulated to determine the complete correcting barcode.

[00335] In some embodiments, three additional hybridization can correct for majority of the errors due to drop or loss of three rounds of hybridization. For example, for 6 rounds of hybridization with 5 types of detectable signals (F=5), three extra rounds of hybridizations are added to create the full 9-member error correction code:

$$\{i, j, k, l, m, n, (k + i - l + m - n)\ mod\ F, (i - l + j - k + m)\ mod\ F, (l - n - j - k + i)\ mod\ F\} \quad (6)$$

[00336] Similar to the previous example, 3 additional rounds of hybridizations can correct for a majority of the loss or drop of 3 rounds of hybridization. There are a total of 84 combinations how 3 rounds of hybridization can be lost or dropped. A 9-componenet code as illustrated in (6) can correct for 72 out of the 84 combinations.

[00337] In some embodiments, 4 additional rounds of hybridizations can correct for the drop or loss of all and any three rounds of hybridization. An example 10-component code is as follows:

$$\{i, j, k, l, m, n, (k+i+l+m+n)\ mod\ F, (i-l+j-k+m)\ mod\ F,\ (l-n-j-k+i)\ mod\ F, (n-k-i-j+m)\ mod\ F\} \quad (7)$$

[00338] It will be understood that there are many other solutions that can be determined empirically. For higher number of drops, similar correction schemes can be determined empirically.

[00339] For 16,000 species, this scheme allows 10 hybs with the ability to correct 3 drops. In comparison, in MERFISH, 16 hybs are needed to target 140 species, with only 2 round correction ability. Because the more round of hybridization one implements, the more mistakes can be made, keeping the number of hybs low is crucial. Thus, this error correction scheme is very powerful compared to the Hamming Distance scheme used in MERFISH. This is because hamming distance correction is used in telecommunications with binary numbers, which uses much longer strings of 0,1.

[00340] As described above, the design disclosed above can correct for loss of 1 hybridization for an arbitrarily long barcode sequence with minimal extra effort. In this example, only one round of error correction is needed in a total of 4 rounds of hybridization that analyzes 100 genes, which below the capacity of 54 (625).

[00341] For example, 7 rounds of hybridization with 5 colors can cover 5^7= 78,125 transcripts, more than the transcriptome, with 8 hybridizations the entire transcriptome can be coded with error correction using the barcoding system disclosed herein.

[00342] Another consideration in designing error-tolerant barcodes is that the mechanism of re-hybridization should guide the robustness of error correction. In the merFISH implementation of seqFISH (Chen 2015), null signal, or "0", along with "1" which is cy5 fluorescence, is used to form a binary barcode. However, it is difficult to determine whether no signal is due to mis-hybridization or actual null signal. In the seqFISH implementation using positive signals as readouts during each round of hybridization reduces the need for error correction because false positive signal is unlikely to re-occur in the same position during another hybridization due to DNAse stripping between hybridizations. Thus, implementation of

seqFISH with 5 colors and 1 extra round of hybridization to error correct is both efficient and accurate, and allows imaging of a large tissue sections since imaging time is ultimately limiting in multiplexing experiments.

[00343] At step 3050, sequential hybridization is carried out to associate or assign barcodes from step 3040 to target genes in a sample. As disclosed herein, the sample can be immobilized mRNAs, DNAs, chromosomal DNAs, and combinations thereof. For example, in the 100-gene sequential hybridization example (see Figure 32 and Figure 40), 4 rounds of hybridization are carried out using probes associated with 5 different types of visual signals. Barcodes are assigned through selection of probes during the 4 rounds of hybridization experiment on immobilized nucleic acid samples.

[00344] At step 3060, after hybridization, visual signals are collected and used in further analysis. For example, images are collected from different hybridization are used to readout the barcodes for specific locations on the immobilized nucleic acid samples. Such barcodes can then be used to decipher the identity of the nucleic acid targets (see, for example, Figures 2, 32, 33, 40 and 41).

[00345] In one aspect, sequential hybridization and serial hybridization are combined for gene identification. In serial hybridization, only one round of hybridization is used to identify target genes. The method is particularly helpful when analyzing genes whose expression level is too high. In some embodiments, genes that are highly expressed, if included in hybridization analysis with genes that are not so highly expressed, would overpower the signals for the genes that are not so highly expression. In some embodiment, the method can also applied to genes whose expression level is too low.

[00346] In some embodiments, expression levels of genes are pre-determined. For example, gene expression levels (e.g., measured by mRNA transcription level) can be already available for certain species. It is possible to identify highly expressed genes by mining publically available data, thus obviating the need to conduct additional experiments to measure expression level.

[00347] In some embodiments, initial experiments are performed to determine relative expression level of candidate genes. In some embodiments, genes are grouped according to their expression levels. For example, genes with moderate or low expression levels can be grouped

together and subject to sequential hybridization analysis. Genes that are highly expressed can be subject to serial hybridization analysis. In some embodiments, expression levels of different genes are compared to the same control gene to derive a relative expression level. For example, the expression level of actin can be used as a control. It will be understood that gene expression level may vary by organisms and can change with respect to different internal and environmental controls. In some embodiments, data from existing expression analysis can be used in identifying highly expressed gene. In some embodiments, preliminary expression analysis is carried out before sequential and/or serial hybridization analysis.

[00348] In some embodiments, a threshold value is set for high expression. Any genes having expression level above the threshold will be excluded from sequential hybridization.

[00349] Depending on types of detectable visual signals that are available, a serial hybridization experiment can detect as many target genes as the number of types of detectable visual signals. For example, in the experiment illustrated in Figures 32 and 40, 5 genes are analyzed at the same time during one serial hybridization experiment.

[00350] In some embodiments, when multiple target genes are present in one serial hybridization round, the number of probes that recognize each target gene is selected such that overlapping of signals is minimize or avoided. In some embodiments, the concentration of probes are selected to avoid or minimize overlapping of detectable signals.

Computer System

[00351] Figure 31 depicts a diagram of an example system architecture for implementing the features and processes of the method disclosed herein, in particular the barcode design functionalities with embedded error correct mechanism.

[00352] In one aspect, some embodiments can employ a computer system (such as the computer system 3100) to perform methods in accordance with various embodiments of the invention. An exemplary embodiment of computer system 3100, includes a bus 3102, one or more processors 3112, one or more storage devices 3114, at least an input device 3116, at least an output device 3118, a communication subsystem 3120, working memory 3130 which includes an operating system 3132, device drivers, executable libraries, and/or other code, such as one or more application(s) 3134 (one or more for implementing the methods disclosed herein).

[00353] According to a set of embodiments, some or all of the procedures of such methods are performed by the computer system 3100 in response to processor 3112 executing one or more sequences of one or more instructions (which might be incorporated into operating system 3132 and/or other code, such as an application program 3134) contained in working memory 3130. Such instructions can be read into the working memory 3130 from another computer-readable medium, such as one or more of storage device(s) 3114. Merely by way of example, execution of the sequences of instructions contained in working memory 3130 might cause processor(s) 3112 to perform one or more procedures of the methods described herein. Additionally or alternatively, portions of the methods described herein can be executed through specialized hardware. Merely by way of example, a portion of one or more procedures described with respect to the method(s) discussed above, such as method 3000, and methods illustrated in Figures 3B and 30, might be implemented by processor 3112.

[00354] In some embodiments, computer system 3100 can further include (and/or be in communication with) one or more non-transitory storage devices 3114, which can comprise, without limitation, local and/or network accessible storage, and/or can include, without limitation, a disk drive, a drive array, an optical storage device, a solid-state storage device, such as a random access memory ("RAM"), and/or a read-only memory ("ROM"), which can be programmable, flash-updateable, and/or the like. Such storage devices can be configured to implement any appropriate data stores, including without limitation, various file systems, database structures, and/or the like. In some embodiments, the storage device 3114 can be example of local database of a user device, or the server database of a server.

[00355] In some embodiments, computer system 3100 can further include one or more input devices 3116, which can comprise, without limitation, any input device that allows a computer device to receive commands from a user such as a request for barcode design.

[00356] In some embodiments, computer system 3100 can further include one or more input output devices 3118, which can comprise, without limitation, any output device that can receive information from a computer device and communicate such information to a user, to another computer device, to the environment of the computer device, or to a functional component communicably connected with the computer device. Examples of input devices include but are not limited to a display, a keyboard, a mouse and etc. For example, the results of

the barcoding analysis can be presented on any one or more of the output devices, for example, as two dimensional heat map, cluster map, a list, a table, and etc.

[00357] It would be understood that any applicable input/output devices or components, such as those disclosed in connection with one or more user device or server, can be applied to input device 3116 and output device 3118.

[00358] In some embodiments, computer system 3100 might also include a communications subsystem 3120, which can include without limitation a modem, a network card (wireless or wired), an infrared communication device, a wireless communication device, and/or a chipset (such as a Bluetooth.TM. device, an 802.11 device, a WiFi device, a WiMax device, cellular communication facilities, etc.), and/or the like. Communications subsystem 620 can include one or more input and/or output communication interfaces to permit data to be exchanged with a network, other computer systems, and/or any other electrical devices/peripherals. In many embodiments, computer system 3100 will further comprise a working memory 3130, which can include a RAM or ROM device, as described above.

[00359] In some embodiments, computer system 3100 also can comprise software elements, shown as being currently located within the working memory 3130, including an operating system 3132, device drivers, executable libraries, and/or other code, such as one or more application(s) 3134, which can comprise computer programs provided by various embodiments, and/or can be designed to implement methods, and/or configure systems, provided by other embodiments, as described herein. Merely by way of example, a portion of one or more procedures described with respect to the method(s) discussed above, such as the methods described in relation to Figure 30, can be implemented as code and/or instructions executable by a computer (and/or a processing unit within a computer); in an aspect, then, such code and/or instructions can be used to configure. and/or adapt a general purpose computer (or other device) to perform one or more operations in accordance with the described methods.

[00360] A set of these instructions and/or code might be stored on a non-transitory computer-readable storage medium, such as storage device(s) 3114 described above. In some cases, the storage medium might be incorporated within a computer system, such as computer system 3100. In other embodiments, the storage medium might be separate from a computer system (e.g., a removable medium, such as an optical disc), and/or provided in an installation

package, such that the storage medium can be used to program, configure, and/or adapt a general purpose computer with the instructions/code stored thereon. These instructions might take the form of executable code, which is executable by computer system 3100 and/or might take the form of source and/or installable code, which, upon compilation and/or installation on the computer system 3100 (e.g., using any of a variety of generally available compilers, installation programs, compression/decompression utilities, etc.), then takes the form of executable code.

[00361] It will be apparent to those skilled in the art that substantial variations can be made in accordance with specific requirements. For example, customized hardware might also be used, and/or particular elements might be implemented in hardware, software (including portable software, such as applets, etc.), or both. Further, connection to other computing devices such as network input/output devices can be employed.

[00362] The terms "machine-readable medium" and "computer-readable medium," as used herein, refer to any medium that participates in providing data that causes a machine to operate in a specific fashion. In an embodiment implemented using computer system 3100, various computer-readable media might be involved in providing instructions/code to processor(s) 3112 for execution and/or might be used to store and/or carry such instructions/code. In many implementations, a computer-readable medium is a physical and/or tangible storage medium. Such a medium can take the form of a non-volatile media or volatile media. Non-volatile media include, for example, optical and/or magnetic disks, such as storage device(s) 3114. Volatile media include, without limitation, dynamic memory, such as working memory 3130.

[00363] Common forms of physical and/or tangible computer-readable media include, for example, a floppy disk, a flexible disk, hard disk, magnetic tape, or any other magnetic medium, a CD-ROM, any other optical medium, any other physical medium with patterns of holes, a RAM, a PROM, EPROM, a FLASH-EPROM, any other memory chip or cartridge, or any other medium from which a computer can read instructions and/or code.

[00364] Various forms of computer-readable media can be involved in carrying one or more sequences of one or more instructions to processor(s) 3112 for execution. Merely by way of example, the instructions can initially be carried on a magnetic disk and/or optical disc of a remote computer. A remote computer might load the instructions into its dynamic memory and

send the instructions as signals over a transmission medium to be received and/or executed by computer system 3100.

[00365] Communications subsystem 3120 (and/or components thereof) generally will receive signals, and bus 3102 then might carry the signals (and/or the data, instructions, etc. carried by the signals) to working memory 3130, from which processor(s) 3112 retrieves and executes the instructions. The instructions received by working memory 1330 can optionally be stored on non-transitory storage device 3114 either before or after execution by processor(s) 3112.

[00366] The methods and systems are provided by way of illustration only. They should in no way limit the scope of the present invention.

[00367] Having described the invention in detail, it will be apparent that modifications, variations, and equivalent embodiments are possible without departing the scope of the invention defined in the appended claims. Furthermore, it should be appreciated that all examples in the present disclosure are provided as non-limiting examples.

[00368] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

EXEMPLIFICATION

[00369] The foregoing has been a description of certain non–limiting embodiments of the invention. Accordingly, it is to be understood that the embodiments of the invention herein described are merely illustrative of the application of the principles of the invention. Reference herein to details of the illustrated embodiments is not intended to limit the scope of the claims.

EXAMPLE 1

Pseudo-color Based Barcoding

Sample preparation

[00370] Figure 5 illustrates mRNA transcripts immobilized on a surface through poly-A-tail or hydrogel embedment. Once cells are lysed, the cell lysate or purified total RNA can be immobilized on a surface through a coverslips functionalized by DNA or LNA Poly-T by

capturing the poly-A tail of mRNA. Alternatively, the mRNA can be mixed with hydrogel such as polyacrylamide gel and allowed gelation on a coverslips surface which the pore size formed would trap the mRNA molecules on the surface.

Design of primary probes

[00371] Figure 6 illustrates an exemplary embodiment of gene specific primary probes design and sequential barcoding hybridization on mRNA immobilized on a surface. In particular, Figure 6a shows that primary probes are not directly labeled with fluorophore. The primary probe contains (i) a gene specific targeting region of the mRNA which is 20nt to 35nt, (ii) one or multiple secondary readout probe binding sites, (iii) PCR primer pairs, and optionally (iv) a restriction enzyme cutting sites and (v) spacers nucleotides. Each target mRNA requires at least 20 or more primary probes. These primary probes are synthesized from a complex oligo pool which later is amplified by the unique PCR primers to obtain a complete set of probes. The primary probes can be attached to target sequence either on the surface or in solution (Figure 6b). Different primary binding sequences are associated with different readout sequences, which can be selectively turned on during hybridization rounds.

[00372] In this case, mRNA 1 is barcoded with red color and mRNA 2 is barcoded with blue color in the first round of hybridization. After imaging, a signal extinguishing step is performed, followed by the next round of barcoding hybridization. Each mRNA will receive 4 colors in total which gives their unique identity to differentiate between each other. The barcode capacity scales up as F^N, with F is the number of fluorophore and N is the number of barcoding hybridization (Figure 6c).

Arrangement of secondary probe binding sites

[00373] Figure 7 illustrates different arrangements of secondary probe binding sites on primary probes. Assume 24 primary probes are used for each gene, 4 rounds of barcoding hybridization with 4 different unique secondary probe binding sequences, there are various combinations of arranging the secondary probe binding sequences on the primary probe. Multiple or all unique secondary probe binding sequences can be found in one probe; for

example, as overhang sequences connected to the primary binding sequence in various combinations (Figure 7a-7c).

Signal Extinguishing Steps

[00374] Figure 8 depicts possible ways to extinguish fluorescent signals. For example either high concentration of formamide can be introduced to 'melt' or strip off the secondary readout probe (Figure 8a). Alternatively, chemical cleavage can be applied (Figure 8b); for example, one implementation can be disulfide conjugated dye on secondary readout probes can be reduced by TCEP or DTT and other reducing agents to cleave off the linker, and thus getting rid the fluorophore from the primary probe..

Scaling up the number of 'fluorophore' by serial hybridization

[00375] Figure 10 depicts a table showing possible barcodes obtained using the pseudo-color barcode scheme. In this example, 3 different color signals, 4 serial hybridization rounds and 4 barcoding rounds were used. In this implementation, 16 rounds of imaging will be implemented with a total of 48 unique secondary readout adapters conjugated to one of the 3 fluorophores.

[00376] One can adjust F, n, and N to achieve the best strategy to generate a desired barcode capacity. For example, in order to reduce the number of barcoding rounds to 3, 30 unique secondary readout sequences are needed at each round of barcoding, as 30^3=27,000. This scheme can allow lower density of targets detected per channel in each hybridization round, allowing a larger number of genes to be detected in situ.

Time efficient error correction scheme

[00377] A time efficient error correction scheme can be implemented in the current coding scheme to tolerate 1 drop of 'color' in decoding the whole color codes due to loss of hybridization. It is possible to design a highly efficient correction schemes that are not perfect by obtaining the tightest packing density for the n-spheres. A barcode generator (i, (i+j+k) mod 5,j,k) can be used to generate barcodes that can tolerate 1 round of mis-hybridization. For example, 6 rounds of hybridization with 5 color can cover 15625 genes, one can design the

barcode scheme to have an extra round of hybridization to still be able to identify the 15625 genes even though any one round of hybridizations is lost in the barcode. More rounds can be corrected with additional error correction hybridization rounds. For example, barcodes can be designed such that shortened barcodes are still unique after data for two or more rounds of hybridizations are lost.

Hybridization experiment

[00378] To show the entire idea works, we targeted 64 genes in mouse NIH3T3 cells (Figure 10). These 64genes were chosen randomly with a range of FPKM values from bulk RNA-sequencing data. A set of 24 primary probes for 1 gene were designed as illustrated in Figure2 which contain 35nt of gene specific targeting region, 3 unique secondary readout binding sequences, a pair of PCR primers, and a 1nucleotide-'T' spacer. The primary probes were synthesized in a complex oligo pool and were amplified by PCR. The barcodes were generated as $4^3 = 64$, without implementing an error correction scheme. 3 fluorophores were used: Cy3b, Alexa594, and Alexa 647 and the number of fluorophores were scaled up as described in Figure 4 with slight variation. Briefly, in each round of barcoding hybridization, 4 unique secondary readout probes were coupled to the 3 fluorophores + one of the fluorophores hybridized in a later hybridization. In this experiment, the fourth unique secondary readout probes in each barcoding hybridization were coupled to Alexa 647, Alexa594, and Cy3B so that only one extra round of imaging is needed to scale up the number of 'fluorophores'. The length of readout probes used in this proof-of-concept experiment is 18mers. These secondary readout probes were conjugated to dye through an amine-NHS reaction. The fluorescent signals in each hybridization was extinguished by high concentration of formamide (60% formamide).

[00379] This proof-of-principle shows that the secondary probe readout scheme using a combination of serial and barcoding steps can be scaled up to perform whole transcriptome profiling. Probes can be designed to target over 20,000 genes in the transcriptome and can be readout with 4 barcoded rounds of hybridization with 12 based "colors", performed in similar fashion as this 64-gene experiment.

[00380] Figure 11 depicts an exemplary embodiment of data analysis. Pearson's correlation plot between the copy number determined by this technology vs FPKM values

determined from bulk RNA-seq. This shows that the measurements by the sequential hybridization barcoding experiments are highly accurate.

[00381] The barcoding scheme illustrates in Figures 8 and 9 were used for multiplex detection of 1000 different transcription factors mRNAs with in vitro RNA SPOTs. First, NIH3T3 cells were grown on a 6-well plate until 60% to 80% confluency. Then the cells were lysed and purified according to Qiagen Rneasy Mini Kit. Then, 100ng of total RNA in RNA Binding Buffer consists of 1M LiCl, 40mM pH7.5 Tris-HCl, 2mM EDTA, 0.1% Triton X-100, and 20units of SUPERase Inhibitor(ThermoFischer) was captured on a LNA functionalized coverslips for overnight at room temperature. The capturing time and amount of total RNA can be modified to adjust the dots density on a coverslip in one frame of view. Next, primary probes(1nM/probe) which consists of a gene specific binding region, 4 barcoded secondary readouts regions, spacers, and primers pair for a thousand genes were allowed to hybridize to the mRNA target in 30% hybridization buffer made from 10% Dextran Sulfate, 2x SSC, and 30% formamide at 37 °C overnight. Each gene is targeted by a minimum of 24 probes. After washing, RNA SPOTs began with the first round of hybridization by flowing in 10nM readout probes for each color (3 colors in one round of hybridization) and the hybridization was allowed to happen for 30minutes at room temperature. After a brief wash, anti-bleaching buffer was flowed into the flow cell, and the fluorescent signals were imaged. Then, the fluorescent signals were extinguished by reduction of 50mM TCEP in 2x SSC with 0.1% Triton X-100 at room temperature for 15mins. The process was repeated until the last round of hybridization.

[00382] In this implementation, 16 rounds of imaging was implemented with a total of 48 unique secondary readout adapters conjugated to one of the 3 fluorophores. The coding space consists of one round of error correction for up to 1728 different targets. In this experiment, a thousand different mRNAs were encoded with the scheme described in Figure 10. The target mRNAs were chosen from a list of transcription factors (master regulators of gene expression) that are conserved between mouse and humans.

[00383] Figure 12 depicts an exemplary a raw image from a barcoding experiment. The raw image shows the results from one of the hybridization round in the 1000 transcription factor genes experiment. The images of each serial hybridizations were first aligned to obtain the 12 'color' barcodes. Then the dots were decoded based on the error correction barcode scheme.

[00384] Figure 13 depicts exemplary results from data analysis, showing decoded 1000 different RNA counts with RNA SPOTs on coverslip and correlate with bulk RNA-Seq Fragments per Kilobase Million (FPKM) for NIH3T3 cells. 753 genes which have a non-zero value of FPKM and a non-zero value of SPOTs are chosen for this correlation. Note that genes with low FPKM values are well detected by RNA SPOTs.

[00385] Figure 14 depicts exemplary results from data analysis, showing correlation between RNA counts from different field of views(FOV)s. Each dot is a single gene with its x value the total number of transcript counted for that gene in FOV 1-5, the y value the counts for FOV 6-10. A high Pearson's correlation coefficient of 0.9131 is observed. This shows that a few FOV is sufficient to accurately quantify transcript abundance.

Imaging-based translational profiling, the RiboCounter

[00386] Intact ribosomes contain 28S rRNA and 18S rRNA. By probing against these rRNA after the initial RNA SPOTs experiment to decode the mRNA identity, one can estimate the number of ribosomes on each mRNA molecule and infers the translational state of the mRNA molecules. By combining this method and RNA SPOTs, we develop the imaging-based polysome profiling to infer the translational state of each detected mRNA molecules by probing against the 28S and 18S rRNA followed by decoding their identity through RNA SPOTs.

[00387] Experimental Design: Cells were lysed and captured on LNA coverslips. Then, by using fluorescence probes either with direct 20mer smFISH or using primary probes, followed by secondary readouts to target the 18S and/or 28S rRNA before RNA SPOTs, one can obtain the intensity for the ribosomes on each mRNA molecules. The identity of each RNA molecules were decoded by RNA SPOTs as described above.

[00388] Results: Figure 15 depicts exemplary results from translational profiling analysis showing the 28S rRNA median intensity extracted from each mRNA molecules of the thousand RNA species. The 28S rRNA intensity for all the 1000 transcription factor genes were all obtained.

[00389] Figure 16 depicts exemplary results from translational profiling analysis, showing the distribution of 28S rRNA median intensity of the gene Myh9 on each mRNA transcripts. The total number of Myh9 transcripts detected are 1169.

[00390] RNA SPOTs of the transcriptome To implement RNA SPOTs at the transcriptome level, probes were designed targeting the coding regions of 10,212 mRNAs with 28 to 32 probes each gene for a total of 323,156 probes. Probes were stringently designed to avoid off-target and cross hybridization (Supplementary information). The primary probes directly hybridize to the mRNAs captured on a Locked Nucleic Acid(LNA) poly(dT) functionalized coverslip and contains a set of overhang sequences that specifies the barcode unique to each transcript (Figure 4A). A 12 "pseudo-color" based scheme was used such that 4 rounds of barcoding are sufficient to cover the transcriptome (12^4=20,736), with an additional round of error correction to compensate for one drop in any round of barcoding [Shah et al, 2016].

[00391] This design minimizes the barcode length to avoid errors from using long barcodes. The 12 base "pseudo-colors" in each round of barcoding is encoded by a set of 12 readout oligos. Three of the readout oligos were hybridized at a time, imaged in the Cy3b, Alexa 594 and Alexa 647 fluorescence channels, and repeated 4 times to iterate through all 12 readout sequences. After each round of hybridization and imaging, the fluorophores are removed by di-sulfide cleavage and followed by the next round of hybridization. A total of 60 readout oligos in 20 rounds of hybridization were used to decode the 10,212 genes targeted. Every four rounds of hybridization were collapsed onto a single image with 12 pseudo-colors. The barcodes were determined from aligning the 12-color images. A common sequence is present in all primary probes and targeted by an oligo labeled with Alexa488 to serve as an alignment marker through all 20 rounds of hybridization. The switching and rehybridization time is fast, with the overall speed limited by imaging speed. Typically, 100 fields of view containing 106 mRNAs can be imaged with 20 rounds of hybridization in a 14-hour period through an automated fluidics system.

[00392] To determine the accuracy of the transcriptome level measurements, we compare the decoded RNA SPOTs data with RNAseq data in fibroblasts and mESCs, and found that they correlated with R=0.86 and R=0.9 respectively. In addition, RNA SPOTs correlated with the gold standard smFISH quantitation with a correlation of R=0.86 in mESCs (23 genes) [Singer 2014] and R= 0.88 in fibroblasts (7 genes). Between two replicates of RNA SPOTs in fibroblasts, the results agree with R=0.94, indicating that RNA SPOTs is a highly robust and

reproducible measurement method. Lastly, comparing genes that were differentially expressed in fibroblasts versus mESCs, we observed the same trend as those detected by RNAseq. For example, pluripotency factors such as Rex1, Esrrb and Sox2 are highly expressed in mESCs but not expressed in fibroblasts as determined by RNA SPOTs (Figure 19), similar to the differences observed by RNAseq.

[00393] Several genes were observed to be outliers in the RNA SPOTs to RNAseq comparison (Figure 19). When we examined the genes in cells directly by single molecule FISH, we observed that their expression levels matched those determined with RNA SPOTs. This indicates that either hybridization efficiency on these genes were low or they were over-detected by RNAseq.

[00394] An advantage of RNA SPOTs is that specific sets of genes can be profiled selectively simply by designing probe sets targeting those genes. In this fashion, ribosomal RNA and highly expressed housekeeping genes can be avoided simply by eliminating those probes from the gene set. As each dot detected in our assay corresponds to a single mRNA, RNA SPOTs is more efficient in term of imaging compared to RNAseq, where many sequencing reads are needed to determine the abundance of a transcript.

[00395] RNA SPOTs can be scaled down to single cell in combination with microfluidics tools to trap and lyse cells [Bose 2015] or with split-pool molecular indexing methods [Cao 2017, Rosenberg 2017]. With targeted RNA SPOTs, we can choose to probe only for the 2000 transcription factors [Fulton 2009] or 1000 landmark informative genes [Donner 2012, Duan 2016] in single cells, instead of profiling the transcriptome, to capture the essential information in cells and to increase the number of cells sampled.

[00396] RNA SPOTs can enable many additional methods beyond expression profiling to study RNA variants [Levesque 2013], modifications [Mellis 2017], RNA binding proteins [Buenrostro 2014] and profile ribosomes occupancy [Ingolia 2009] directly on the captured RNA. Lastly, noncoding RNAs and other RNAs without polyA tails can be captured in hydrogels rather than with dT oligos. As cost of sequencing is a major limiting factor in genomics experiments, SPOTs enable an accurate and low-cost alternative to sequencing with many further applications beyond RNA to DNA and proteins.

EXAMPLE 2

Readout Probes and Rehybridization in Mouse Embryonic Stem Cells

Synthesis of DNA Probes-Disulfide-Dye Conjugates

[00397] An exemplary scheme for synthesizing readout probes-dye conjugates connected by a disulfide bond. Thiol-modified DNA probes were ordered from Integrated DNA Technologies in their oxidized form. 10nmoles of thiol-modified DNA probes was treated with 10mM TCEP at 37°C for 30 minutes. After reduction step and gel column purified, the DNA probes were mixed with 50 equivalents of 3-(2-Pyridyldithio) propionic acid N-hydroxysuccinimide ester (SPDP) linker in 1x PBS solution containing 10mM EDTA. The mixture was allowed to react at room temperature for at least 2hours. Immediately after the reaction, the mixture was spin column purified and was re-suspended in 60uL of 1x PBS containing 100ug of cadaverine dyes. The reaction was allowed to proceed at room temperature for at least 4 hours before subjected to ethanol-precipitation purification and HPLC purified. The concentration of the final product was determined using Nanodrop.

Rehybridization in mouse embryonic stem cells (mESCs).

[00398] Figure 28 illustrates rehybridization reactions in mouse embryonic stem cells (mESCs). To verify the scheme works in both cells and on coverslips, we perform a proof of concept experiments in mouse embryonic stem cells (mESCs) targeting introns of pgk1 gene. During first round of hybridization, tertiary readout probes conjugated with A647 were applied to identify specific target sequences in mRNAs via secondary bridge binding sites. Real fluorescent spots are shown in red dashed box (28A). As a control, fluorescence presence in channel 594 was also measured during the first hybridization: no fluorescent spots were observed (28B). After TCEP treatment and washing steps, channel 647 was reimaged. The observed dim dots are non-specific binding of dyes which does not interfere with subsequent real fluorescent spots identification (28C). During a second round of hybridization, unique readout probes bearing 594 dye were used to hybridize with the secondary bridge binding sites to give real fluorescent spots (28D). These spots confirmed the positions observed previous when readout probes with A647 were used (28A).

In vitro rehybridization with readout probes

[00399] Transcribed polyA-tailed dCAS9-EGFP mRNA were captured on a dT20 Locked Nucleic Acid (LNA) surface-modified coverslips to show the rehybridization scheme works *in*

vitro on the coverslips. Once again, specific binding was observed. When tertiary probes conjugated with A647 were used: specific signals were observed in channel 647 while there were few signals observed in channel 594 (Figure 29, top row). After TCEP treatment and wash of the first set of tertiary probes, a second set of tertiary probes conjugated with 594 dye were applied during the set round of hybridization reaction. Consequently, specific signals were observed in channel 594 while there were few signals observed in channel 647 (Figure 29, bottom row).

[00400] Technically, any heterobifunctional cross-linking reagent that can connect between the dye and thiol-modified DNA probes will work for this rehybridization scheme. DNA probes-disulfide-dye conjugates were synthesized using 3-(2-pyridyldithio) propionyl hydrazide (PDPH) linker and NHS ester dyes which work equally well as former conjugates.

EXAMPLE 3

Brain Slice Analysis

[00401] As an illustration, barcodes generated using the error correction mechanisms disclosed herein are used for *in situ* transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus.

[00402] Identifying the spatial organization of tissues at cellular resolution from single cell gene expression profiles is essential to understanding many biological systems. In particular, there exist conflicting evidence on whether the hippocampus is organized into transcriptionally distinct subregions. Here, a generalizable *in situ* 3D multiplexed imaging method was applied to quantify hundreds of genes with single cell resolution via Sequential barcoded Fluorescence *in situ* hybridization (seqFISH) (Lubeck et al., 2014). seqFISH was used to identify unique transcriptional states by quantifying and clustering up to 249 genes in 16,958 cells. By visualizing these clustered cells *in situ*, we identified distinct layers in the dentate gyrus corresponding to the granule cell layer, composed of predominantly a single cell class, and the subgranular zone, which contains cells involved in adult neurogenesis. Furthermore, it was discovered that distinct subregions within the CA1 and CA3 are composed of unique combinations of cells in different transcriptional states, instead of a single state in each sub-region as previously proposed. In addition, it was revealed that while the dorsal region of the CA1 is relatively homogenous at the single cell level, the ventral part of the CA1 has a high

degree of cellular heterogeneity. These structures and patterns are observed in sections from different mice, as well as in seqFISH experiments with different sets of genes. Together, these results demonstrate the power of seqFISH in transcriptional profiling of complex tissues.

[00403] The mouse brain contains about 10^8 cells arranged into distinct anatomical structures. While cells in these complex structures have been traditionally classified by morphology and electrophysiology, their characterization has been recently aided by gene expression studies. In particular, the Allen Brain Atlas (ABA) provides a systematic gene expression database using in situ hybridization (ISH) of the entire mouse brain one gene at a time (Dong et al., 2009; Fanselow and Dong, 2010; Thompson et al., 2008). This comprehensive reference provides regional gene expression information, but lacks the ability to correlate the expression of different genes in the same cell. More recently, single cell RNA sequencing (RNA-seq) has identified many cell types based on gene expression profiles (Darmanis et al., 2015; Tasic et al., 2016; Zeisel et al., 2015). However, while single cell RNA-seq provides useful information on multiple genes in individual cells, it has relatively low detection efficiencies and requires cells to be removed from their native environment resulting in the loss of spatial information. These different approaches can lead to contradictory descriptions of cellular organization in the brain and other biological systems.

[00404] In the hippocampus, recent RNA-seq data suggests that CA1 is composed of cells with a continuum of expression states (Cembrowski et al., 2016, Zeisel et al 2015), while ABA analysis indicates that sub-regions within the CA1 have distinct expression profiles (Thompson et al, 2008). To resolve the two conflicting descriptions of hippocampal organization, a method to profile transcription *in situ* in the hippocampus with single cell resolution is needed. Here, we demonstrate a general method that enables the mapping of cells and their transcription profiles with single molecule resolution in tissue, allowing an unprecedented resolution of cellular transcription states for molecular neuroscience (Figure 32A).

[00405] A great deal of progress has been made recently in developing highly quantitative methods to profile the transcriptome of single cells. Building upon single molecule fluorescence *in situ* hybridization (smFISH) (Femino et al., 1998; Raj et al., 2006;) , Lubeck et al. devised a general method to highly multiplex single molecule *in situ* mRNA imaging irrespective of transcript density using super-resolution microscopy (Betzig et al., 2006; Rust et al., 2006;

Lubeck and Cai, 2012). However, the spectral barcoding methods used in these previous works is difficult to scale up beyond 20-30 genes because of limited number of fluorophores (Fan et al., 2001; Lubeck and Cai, 2012).

[00406] To overcome the scalability problem, a temporal barcoding scheme was developed that uses a limited set of fluorophores and scales exponentially with time (Lubeck et al., 2014). Specifically, by using sequential rounds of probe hybridizations on the mRNAs in fixed cells to impart a unique pre-defined temporal sequence of colors, different mRNAs can be uniquely identified *in situ*. The multiplex capacity scales as F^N, where F is the number of fluorophores and N is the number of rounds of hybridization. Thus, one can increase the multiplex capacity by increasing the number of rounds of hybridization with a limited pool of fluorophores. This approach is called Sequential barcoded Fluorescence i*n situ* Hybridization (seqFISH) (Lubeck et al., 2014). In parallel, *in situ* sequencing methods were developed to directly sequence transcripts in tissue sections, but these methods suffer from low detection efficiency (<1%) (Ke et al., 2013; Lee et al., 2014). Recently, Chen et al. expanded the error correction method in the original seqFISH demonstration by using a Hamming distance 2 based error correcting barcode system, called merFISH. However, this implementation requires larger transcripts (>6kb) and many more rounds of hybridization than the method described here (Chen et al., 2015b). Furthermore, seqFISH and its variants have only been applied in cell culture systems due to the difficulty of smFISH detection in tissue. Here, an improved version of seqFISH in complex tissues by including signal amplification and a time-efficient error correction scheme (Figures 32A-D, Figure 40) were demonstrated to resolve the structural organization of the hippocampus with single cell resolution.

EXAMPLE 4

Brain Slice Analysis with Error Correction

[00407] *Signal amplification and error correction enable robust detection of mRNAs in tissues.*

[00408] To overcome the autofluorescence and scattering inherent to brain tissues, we used an amplified version of smFISH, called single molecule Hybridization Chain Reaction (smHCR) (Figure 32E) (Shah et al., 2016). Single molecule HCR amplified signal 22.1 ± 11.5 (mean ± s.d., n=1288, Figure 38B) fold compared to smFISH, enabling robust and rapid

detection of individual mRNA molecules in tissues and facile alignment of spots between hybridizations (Figure 38A). Single transcripts can be detected and localized in 3D with just 24 probes in tissues, enabling detection of transcripts <1kb in size,with a fidelity comparable to the smFISH gold standard (Figures 41C-D) but with signals 20-fold brighter (Shah et al., 2016). Single molecule HCR DNA polymers can also be digested by DNAse and re-hybridized in brain slices, allowing HCR-seqFISH to be robustly implemented (Figure 33A). We note the smHCR enables true 3D imaging in tissues, whereas the previous sequential FISH demonstrations (Lubeck et al., 2014, Chen et al., 2015) were performed only in flat cell cultures.

[00409] Furthermore, we improved upon our existing barcode system by implementing a time-efficient error correction scheme. The major source of error in seqFISH is the loss of signal due to mis-hybridization, which increases with the numbers of hybridization. We introduced an extra round of hybridization to correct loss of signal during any round of hybridization (Figure 32D). By minimizing the number of hybridizations, this error correction scheme is efficient in practical implementation. For example, using 5 fluorophores and 4 rounds (instead of 3 rounds) of hybridization to code for 125 genes, we can still uniquely assign barcodes to genes even when signal from any single round of hybridization is missing. Although merFISH can tolerate 2 errors in the barcodes, it requires 16 rounds of hybridization to code 140 genes (Chen et al. 2015). As increasing the number of hybridizations can potentially lead to more experimental error and analysis complexity, our simple error correction method corrects for the most common error, dropped signal. Also, the fewer rounds of hybridizations decreases the total imaging time, which is rate-limiting for tissue experiments. HCR-seqFISH with simpler error-correction scheme allows efficient and accurate quantification of transcription profiles in tissues.

[00410] Using this HCR-seqFISH method, we surveyed the regional and sub-regional transcriptional heterogeneity within the temporal and parietal cortex and hippocampus of the mouse brain by imaging similar coronal sections collected from 3 different animals. Two similar sections from separate mice were profiled with probes for 125 genes, while one additional brain slice was imaged for 249 genes. In each of the coronal slices, between 60-80 fields of view were imaged, each 216μm x 216μm x 15μm, in the cortex and hippocampus (Figure 32A and Figure 41E). For the 125 gene set, 56 of the genes (Figure 32D, Figure 40) were selected because they showed spatially heterogeneous expression based on the ABA (Lein et al., 2007), another 44

were selected from a list of transcription factors, and 25 marker genes were selected from single cell RNA-seq datasets (Zeisel et al., 2015). One hundred of these genes were barcoded by 4 rounds of hybridization (Figure 32B). The remaining 25 high abundance genes were measured individually using 5-color smHCR in 5 serial rounds of hybridizations (Figure 32C). This hybrid approach of measuring medium expression genes with barcoding seqFISH and high copy number genes serially in subsequent hybridizations allows a large dynamic range of transcripts to be profiled in the same cell.

[00411] *seqFISH is an accurate and efficient method to multiplex RNA in situ.*

[00412] To determine the accuracy of the seqFISH method in quantifying mRNA levels in single cells in tissue, we compared the copy number of 5 of the 100 target genes measured by barcoding to the copy number found by colocalized smHCR detection in the same cell (Figure 33B, Figure 42A) in 15μm brain sections. We found that the copy number of the RNAs per cell as measured by barcoding and smHCR agreed with an R-value of 0.85 and a slope of 0.84 (N=3851). As colocalized smHCR matches smFISH transcript quantitation (Shah et al., 2016), the barcoded seqFISH method can quantify mRNA molecules in single cells with 84% efficiency compared to the gold standard of smFISH. In comparison, single cell RNA-seq measurements are 5-20% efficient based on spike-in controls and *in situ* sequencing is less than 1% efficient (Darmanis et al., 2015; Klein et al., 2015; Lee et al., 2014; Macosko et al., 2015; Tasic et al., 2016; Zeisel et al., 2015; Ståhl et al., 2016). This high efficiency of detection results from a low transcript drop rate and a high barcode recovery rate due to the error correction round of hybridization. In our experiment, 78.9% of barcodes (N=2,115,477 barcodes) were found in all 4 hybridization rounds and 21.1% were identified in 3 out of the 4 hybridizations (Figure 33C), indicating that the probability of detecting a given mRNA molecule is 94% in each round of hybridization (Figure 42B).

[00413] To quantify the amount of false positive signal due to misalignment of barcodes and nonspecific binding of probes, the amount of off-target barcodes that were detected was measured. With four rounds of hybridizations and 5 fluorophores, there were 5^4=625 unique codes. 100 of these barcodes were assigned to measure mRNAs detected at 914.8 ± 570.5 counts per cell (mean ± s.d., N=3439). In comparison, the 525 remaining off-target barcodes that were not used were detected at 4.6 ± 4.7 (mean ± s.d., N =3439) counts per cell (Figure 33D). False

positives, due to chance alignment of nonspecifically bound spots, contributed minimally to the barcode readouts because of this three order of magnitude difference in detected barcodes (on target vs. off target). The false positives we observe fall only on barcodes hamming distance one away from on-target barcodes, yet minimally contribute to undercounting on-target barcodes (Figure 33E). Furthermore, even the most frequent off-target barcode was observed 65.57 times less frequently than the most infrequent mRNA coding barcode (Figure 33E, Figure 42). Even though during each round of hybridization, 24.8 ± 0.4% (mean ± s.e., N=4 rounds of hybridization) of the spots were nonspecifically bound probes, barcode miss-assignments did not occur frequently because non-specifically bound probes do not reappear in the same location after digestion with DNAse and re-hybridization (Figure 33A). Together the quantifications of false positive and false negative barcodes demonstrate that this method is highly efficient and accurate at detecting RNAs *in situ* in single cells within tissues.

[00414] *Cell clusters are based on combinatorial expression profiles.*

[00415] We imaged the expression of 125 genes in coronal sections from two mice for a total of 14,908 cells (Figure 41E). Cortical and hippocampal cells were segmented based on DAPI and Nissl staining. A tessellation algorithm was developed to accurately segment densely packed cells in the hippocampus. To avoid capturing mRNA from neighboring cells, we contracted by 10% the borders of cells determined by the segmentation algorithm.

[00416] To group the single cell data into distinct transcriptional states, we Z-score normalized the copy number of each transcript in every cell (Figure 34A) and hierarchically clustered the cells to identify cells with similar expression patterns (Figure 43). Many of these clusters, based on overall expression patterns, contain clear transcriptional markers of known cell types previously identified by single cell RNA-seq (Figure 34B) (Zeisel et al., 2015, Tasic et al 2016). Cell clusters 12 and 13 contained clear expression of *Gja1* which marks out astrocytes (Zeisel et al., 2015, Tasic et al 2016). Cluster 12 also expresses *Mfge8* while cluster 13 did not, indicating two distinct population of astrocytes (Figure 34B). There are further subclusters within each of the astrocyte populations with different spatial localization patterns (Figure 43C). Cluster 11 cells expressed *Laptm5*, a known microglia marker (Zeisel et al., 2015, Tasic et al 2016). Cluster 3 expressed interneuron genes while cluster 1-2 and 4-5 expressed genes associated with pyramidal neurons (Zeisel et al., 2015, Tasic et al 2016). Some clusters

contained many distinct subclusters, such as Amigo2 enriched Mural cells (cluster 9.4) or Omg expressing oligodendrocytes (cluster 10.4 and 10.5). The major clusters were robust to down-sampling the number of cells used in clustering (Figure 44), with some of the hippocampal pyramidal and glia clusters robustly defined even with 400 cells. Similarly, principal component analysis (PCA) visualization of the data (Figure 43F) recapitulated the major clusters that corresponded to astrocyte, microglia, cortical pyramidal, hippocampal pyramidal, dentate gyrus (DG) granule, and interneuron cells.

[00417] *Cell clusters show distinct regional localization*

[00418] Many neuronal clusters mapped to distinct regions in the brain (Figure 34B). Several classes of pyramidal cells (cluster 1-2) showed exclusive localization to the hippocampus, while other classes (4-5) showed predominantly cortical localization. There were also a class of cells (cluster 7) that were almost exclusively present in the DG. Interestingly, these clusters segregated based solely on gene expression profiles without adding any spatial information into the clustering algorithm. These differences in transcriptional states of neurons could be due to intrinsic differences in the cells or due to different local environment and activity patterns.

[00419] In contrast, astrocyte, microglia and other non-neuronal cell clusters were generally uniformly present in all areas of the brain (Figure 34B). However, subclusters of astrocytes did localize to different regions of the brain preferentially (Figure 43C), with subcluster 12.3 localized preferentially to the cortex, while 12.1 subcluster was uniformly distributed. Similarly, cluster 9 cells contain subclusters (9.3, 9.5 and 9.6) that localize exclusively to the DG, while other subcluster (9.1) localize almost exclusively to the cortex. The regional localization of neurons are especially pronounced with cluster 1 and 2 localized almost exclusively to the hippocampus, with some of the subclusters localized predominantly to the CA3. Furthermore, while pyramidal cell clusters 4 and 5 are preferentially cortically localized, the few hippocampal cells in these clusters form their own subclusters (4.4 and 5.4) (Figure 43C). In cluster 6 cells, many subclusters with distinct expression profiles are localized almost exclusively in the CA1, CA3 or the DG (Figures 34C, 43C). In contrast, cluster 7 cells show a relatively homogenous regionalization pattern, but further subdivide based on combinatorial expression patterns (Figure 34D). Subclusters of cluster 9 also show significant regionalization where subclusters 9.1, 9.3, 9.5, and 9.6 show localization to the SGZ (Figure 43C). Overall, cell

clusters with similar expression profiles exhibited similar spatial localizations across the brain with a correlation coefficient of 0.67 (Figure 43E), indicating the existence of archetypal regional expression patterns and potential spatial markers in the brain. These results show that the tissue-optimized HCR seqFISH approach can directly identify a variety of transcriptional states and quantify broad spatial patterns of expression.

[00420] *Combinatorial expression patterns define fine clusters.*

[00421] While certain cell clusters contain strong expression of marker genes, not all clusters are defined based on a few genes. How much power do individual genes or groups of genes have in explaining the observed cell clusters? To understand this, we examined whether subsets of genes can recapitulate the observed clusters (Figure 34E). We found that any set of 25 genes recovers about half of the correlation structure in the cell-to-cell correlation map (Figures 34E, 45B-C, and 44, N=10 bootstrap replicates). The fact that the selection of any 25 genes can explain the gross patterns in the data is likely due to the high correlations amongst the expression patterns of genes, as shown in the gene-to-gene correlation map (Figure 45A). Thus, a small subset of the measured genes can provides sufficient information to infer the gross transcriptional states of the cells. Interestingly, this may be the same reason why low-coverage single cell sequencing methods such as drop-seq and inDrop (Klein et al., 2015; Macosko et al., 2015) can capture the large distinction of cell types, because many highly expressed genes are correlated to other genes that collectively define cell types.

[00422] At the same time, the finer correlation structure in the data, required to define the cell clusters accurately, can only be captured with more genes (Figures 34F, 45B-C). Consistent with this, using a "random-forest" machine learning algorithm (Breiman, 2001) to classify cell clusters, we found that 75 genes are needed to classify cells with 50% accuracy, indicating that correct cluster assignment requires more detailed information from many genes (Figure 34E). Supporting this view, the first 10 principal components (PC) explained 59.5% of the variation in the data, while the rest of the variation required the remaining 115 PCs (Figures 34F, 43D). The "random forest" algorithm required 10 PCs to predict the cell cluster assignments with 50% accuracy (Figure 34F), but accuracy steadily increased with more PCs. These observation indicated two levels of information in the data: a coarse level, where large distinctions in cell

clusters are observable by a few genes, and a fine level, where subtle distinctions require many more genes.

[00423] These results suggest two points experimentally. First, multiplexing at the level of 20 genes by seqFISH can give broad cell cluster identification that is not available with 2-3 gene smFISH experiments. Although single marker genes are useful for inference, we find that they frequently are not sufficient for cell classification. For example, all DG specific granule cells (clusters 7) have *Gpc4* and *Vps13c* as their enriched marker genes (Figure 34B); yet, *Gpc4* and *Vps13c* are also strongly expressed in other hippocampal cells outside of the DG, as seen in both our experiments and the ABA. Thus, smFISH against *Gpc4* and *Vps13c* alone would not be sufficient to uniquely identify the DG granule cells. Furthermore, even the strongly bimodal markers that are known to define cell types (i.e. *Mgfe8*, *Gja1*, etc.) are correlated enough to overall expression profiles that cells fall into the appropriate cluster even when these genes are excluded. This point suggests that while marker genes can be essential in assigning a cell to a known cell type, they are not necessary to identify unique clusters in the dataset provided enough measurements are made. Second, accurate measurement of combinatorial expression of many genes enabled by seqFISH can allow for more specific cell cluster identification. As a comparison, in single cell RNAseq data, CA1 pyramidal cells are clustered into a single cluster (Zeisel et. al, 2015; Habib et. al 2016) potentially because of the relatively lower detection efficiency of the method. In our seqFISH experiments, measuring hundreds of genes quantitatively, we can resolve several clusters and subclusters with robust regionalization within the CA1 (Figures 34B, 43C).

[00424] *Cells are patterned in the dentate gyrus.*

[00425] To further visualize the spatial organization of cells, we mapped cluster definitions of cells back into the images. In the DG, we observed a striking lamina layering of cell classes. The two blades of the DG (Figures 35A-B) showed mirror arrangements of cells, with cluster 9 cells, forming the subgranular zone (SGZ), leading into a granule cell layer (GCL) dominated by a single cluster of granule cells (cluster 7) (Figure 34B). In the 125 gene data set, the cells of the GCL were found to be dominated by expression of Gpc4 and Vps13c matching ISH data from the ABA (Figure 48B). Cluster 7 was found to be further subdivided into 6 subclusters (Figure 40C). These subclusters were found to have varying levels of calbindin D-

28K (*Calb1*) expression which is known to increase with granule cell maturation (Figure 34D)(Yang et al., 2015). On the other hand, the cells of the SGZ were found to be significantly enriched in astrocyte markers such as *Mfge8* and *Mertk*, which has been also been observed previously (Miller et al, 2013) and in the ABA data. However, these cells do not cluster with typical astrocytes (cluster 12 and 13) because their combinatorial expression patterns are different from astrocytes, consistent with their classification as a completely different population of cells.

[00426] In the fork region of the DG, the layer of cluster 9 cells appeared on the interior surface of the fork, followed by a layer of granule cells (cluster 7) (Figure 35C). A different layering pattern is seen at the crest of the DG, where astrocytes, microglia, and some other glial cells line the exterior of the crest ensheathing the GCL (Figure 35D). In both brains of the 125 gene experiments, the same cell clusters and spatial arrangements are observed. Furthermore, because the mRNAs are imaged in 3D in the 10-15um brain slices, we can obtain a 3D view of the expression profiles, shown in the fork regions of the DG (Figure 35F).

[00427] *Distinct regions of CA1 and CA3 are composed of different combination of cell clusters.*

[00428] While each region of the DG contains similar compositions of cells, distinct subregions within the CA1 and CA3 contained different combinations of cell classes (Figures 36, 46F). In the CA1, there were 3 distinct regions defined by their individual cellular compositions. In the dorsal region of CA1 (CA1d), neuron cluster 6 (enriched in *Nell1*, a protein kinase C binding protein) was the major cell type in the pyramidal layer, with astrocyte, microglia and other cells (clusters 10-13) intercalating into the stratum pyramidale (SP) (Figures 36A-C). Transitioning into the CA1 intermediate region (CA1i) (Figure 36D), pyramidal cell cluster 4 displaced cell cluster 6 as the dominant cell, with the co-appearance of cluster 1 and 2 pyramidal cells.

[00429] As the middle of the CA1i region was reached, a small amount of cluster 4 pyramidal cells remain, while cluster 1 and 2 pyramidal cells dominate (Figures 36E-F). Cluster 1 and 2 are enriched in *Nell1* (EGF like protein), *Npy2r* (neuropeptide Y receptor), *Slc4a8* (sodium bicarbonate transporter) and *B3gat2* (glucuronosyltransferase). The CA1i region displayed a characteristic spatial organization where glial cells line the outermost regions, while

pyramidal cell cluster 1 and 2 longitudinally partitioned the pyramidal layer. This separation of the inner versus the outer layers of CA1 matches those observed in previously (Dong et al., 2008). Furthermore, interneurons (cluster 3) were found to preferentially line the inner edge of the pyramidal layer in the CA1i region (Figures 36E-F). This patterning of interneurons, particularly subcluster 3.1 cells which were enriched in Slc5a7, a choline transporter, was consistent with the patterning of cholinergic interneurons observed with ChAT-GFP labeling (Yi et al., 2015). Finally, the largest amount of heterogeneity in the CA1 was seen in the ventral CA1 region (CA1v), where cell clusters 3, 5, and 10 began to mix in with clusters 1 and 2 (Figures 36G-I).

[00430] Similarly, the CA3 was found to have four transcriptionally distinct regions with different pyramidal cell compositions and abrupt transitions. The ventral most region of CA3 contained a high level of heterogeneity of pyramidal cell clusters (Figures 36J-K), while the intermediate region of CA3 contain a mixture of cell clusters 1 and 2 (Figures 36L-M). As the CA3 progressed towards the hilus of the DG, the cell types transitioned first to primarily cluster 4 neurons (enriched in *dcx*, doublecortin, and *Col5a1*, a collagen), and then to almost exclusively cluster 6 neurons in the region most proximal to the DG hilus (Figures 36O-P). It is interesting to note that while cluster 6 cells appear in both the CA1 (subcluster 6.8) and CA3 (subclusters 6.1 and 6.4), sub-clusters of 6 show distant regional localization (Figure 43C), suggesting that the gene expression differences in CA1 and CA3 cells are captured in the seqFISH data.

[00431] The regionalized expression patterns we observed in the hippocampus match closely to those observed in previous literature (Thompson et al Neuron 2008 and Dong et al PNAS 2009). For example, CA1d, CA1i, CA1v boundaries correspond to the boundaries shown in Fig 2B in Dong et al. In CA3, the subregions observed in our experiment match the CA3 subregion 4-7 in Thompson et al. (Thompson et al., 2008).

[00432] Lastly, we note that the two slices from two different mice in the 125 gene experiment show not only the same subregional structure (Figures 35-37), but also the same clusters of cells (Figures 36 and 37) in the different subregions of the hippocampus (Figure 46). In both brains, the CA1d consists of relatively homogenous population of cluster 6 cells, which transition to a mixture of 1 and 2 cells in CA1i, and finally to a mixture of 1-6 and 10 cells in the CA1v (Figure 46F). These results together show that the sub-regions of the hippocampus are a

robust feature in the organization of CA1 and CA3, consisting of cells classes with distinct expression profiles. The stereotypical nature of the spatial arrangement of these structures suggest further experiments with seqFISH and other functional assays to probe the distinct functions of the different cell clusters in the CA1 and CA3.

[00433] *249 gene multiplex experiments show the same hippocampal subregions*

[00434] To further show that the sub-regional structure of the hippocampus is independent of the target genes, we performed a 249 gene seqFISH experiment on a third coronal section. Of these 249 genes, only 22 genes overlapped with the 125 gene experiment set. For this set of genes, 214 were selected from a list of transcription factors and signaling pathway components and the remaining 35 were selected from cell identity markers from another single cell RNAseq dataset (Tasic et al, 2016). The 214 genes were barcoded by 5 rounds of hybridization, while the remaining genes were imaged in 7 rounds of non-barcoding serial hybridization. To quantify the efficiency of this experiment, 4 genes in the barcoding set (*Smarca4*, *Sin3a*, *Npas3*, and *Neurod4*) were re-probed with smHCR. The barcoding efficiency of the 249 gene probe set was found to be 71% with and R value of 0.80 (Figure 46D). In single cells, we detect on average 2807±1660 (mean±s.d., N=2050 cells) total barcoded barcodes.

[00435] The same arrangement in the DG was observed in the 249 gene experiment, despite different genes used, indicating robust identification of the layering in the DG by seqFISH (Figures 38S-T). In particular, the cells in the SGZ are clustered independently from cells in the GCL, similar to the layers observed in the 125 gene experiment. In the SGZ cells, we observed enrichment of Sox11, a key transcription factor in neurogenesis (Miller et al, 2013). Other transcription factors involved in neurogenesis, NFIA and Tbr1 are also enriched in the SGZ cells as seen in our data and the ABA images (Figure 48A). The observations of this distinct layer in both the 249 and 125 gene experiment and the combined gene enrichment pattern (increased Sox11, Sox9, NFIA, and Tbr1 in the 249 gene experiment and increased Mertk and Mfge8 in the 125 gene experiment) suggests that many cells in this layer are involved in adult neurogenesis in the SGZ.

[00436] In addition, the same regionalized cellular patterns are observed in CA1d, CA1i, and CA1v, where different subregions utilize different cell classes in characteristic ratios (Figure 46F). As seen with the 125 gene experiment, while the CA1d uses only a few cell classes and is

relatively homogeneous, while the CA1v region is made up of many different cell classes resulting in a high level of cellular heterogeneity. Furthermore, the distinction between CA1 and CA3 cell clusters are more clear in the 249 gene experiment suggesting more resolving power of spatial patterns (Figures 38A-K). The 249 gene experiment also suggests that the CA3 may be composed of 3-4 subregions based on cell cluster composition (Figures 38L-R). The cellular heterogeneity of the CA3 is again shown to mirror that of the CA1, where the cellular heterogeneity increases along the dorsal to ventral axis. Cells with distinctive marker gene expression in the hippocampus are shown in supplementary Figure 38A.

[00437] *Single cell data resolves cellular organizations in the sub-regions of the CA1 and CA3.*

[00438] Two conflicting views of the cell types in the hippocampus have been proposed based on the analysis of the Allen Brain Atlas data (Thompson 2008) as well as recent RNA-seq data (Cembrowski et al., 2016, Zeisel et al 2015). Analysis of the ABA in situ data showed that distinct subregions of the hippocampus expressed different molecular markers, indicating that the CA1 and CA3 are "regionalized" into distinct sub-structures (Fanselow and Dong, 2010; Thompson et al., 2008). However, recent bulk RNA-seq experiments on the CA1 found that gene expression patterns changed gradually along the dorsal to ventral axis, contradicting the sharp boundaries observed in the ABA analysis (Cembrowski et al., 2016). Further supporting this "continuous" cell type view of the hippocampus, analysis of the single cell RNA-seq data (Zeisel et al, 2015) identified a single continuous population of cells in the CA1 region.

[00439] Our data provides a single cell resolution picture of the spatial organization of cells in the hippocampus and reconciles both the RNA-seq and the ABA data. While our data mostly supports a regionalized view of the hippocampus, we observe that a single cell class does not in general define CA1 and CA3 sub-regions. Instead, we observed that different subregions of CA1 and CA3 are composed of distinct combinations of cell clusters (Figures 33-35). For example, CA1d consists primarily of cluster 6 pyramidal cells (Figures 36A-C), in addition to the cluster 1,2, 10, and 12cells, while CA1v consists of a large set of cell classes including cluster 1-6 and 10 cells, but at different relative abundances (Figures 36-37, Figures 46F-G). Due to this intermixing of cell classes in each sub-region, a bulk measurement of transcription profiles would find a lack of regionalization, but single cell analysis with spatial resolution would

identify these distinct regions based on their unique cell class compositions. Indeed, when we averaged the single cell expression profile within each sub-region of the CA1, we can reproduce the continuous correlation profiles found by bulk RNA-seq between CA1v, CA1i, and CA1d (Figure 39) (Cembrowski et al., 2016). The bulk RNA-seq observation that CA1i lacked specific marker genes can also be explained. This is in fact consistent with our findings that CA1i contained cell classes present in both CA1d and CA1v (Figures 36-38).

[00440] This organization of cell classes is observed in both the 125 gene experiments as well as in the 249 gene experiment. It is worth noting that the complexity of cell populations observed in the CA1d versus the CA1v matches the functional differences in CA1. CA1d is responsible for spatial learning and navigation and contains a higher concentration of place cells and send projections to dorsal subiculum and cortical retrosplenial area (Cenquizca and Swanson, 2007; Jung et al., 1994; Risold et al, 1997; O'Keefe and Dostrovsky, 1971). We observed that CA1d is composed of a relatively homogeneous population of cells, predominantly of cluster 6 cells. In contrast, the ventral region is involved in a variety of cognitive tasks, such as stress response, emotional and social behavior (Cenquizca and Swanson, 2007; Jung et al., 1994; Fanselow and Dong, 2010; Kishi et al., 2006; Muller et al., 1996; Petrovich et al., 2001; Pitkänen et al., 2000; Saunders et al., 1988; Witter and Amaral, 1991; Yi et al., 2015). Correspondingly, we observed a large set of cell classes in the CA1v regions. It is intriguing to hypothesize that the different cell classes identified based on molecular profiles may correspond to neurons with distinct connectivity and functional patterns. This hypothesis can be investigated in future experiments combining anterograde tracing as well as electrophysiological recording followed by seqFISH.

[00441] A list of the 249 genes being analyzed can be found in the following Table 3.

Name of Genes being analyzed
Tal1
Dmbx1
Emx2
Uncx
Paxip1
Ctnnb1
Prdm1
Rybp
Nfkb2

Name of Genes being analyzed
Tfdp2
Grhl1
Sp8
Irf2
Zfp287
Esr2
Zfp128
Vav1
Sp1
Ppargc1b
Sp7
Pin1
Nfya
Vsx1
Klf1
Vsx2
Mybl1
Mybl2
Rnf2
Blzf1
Topors
Nr3c2
Nfia
Taf6l
Nr4a3
Hoxd12
Hoxd13
Ttf1
Sox9
Nr2e1
Polr2b
Hltf
Sox6
Pbx3
Sox5
Foxa1
Cdc5l
Cebpg
Ciita
Rest
Ets1
Mafk

Name of Genes being analyzed
Tbx15
Scml2
Myb
Clock
Rbpj
Foxc1
Zfp422
Pias3
Runx1
Ppara
Relb
Vdr
Cdc6
Arid3a
Lhx1
Hoxb8
Hoxb9
Hic1
Lhx6
Six4
Hoxb3
Zfp263
Cbfa2t3
Ehf
Nhlh1
Gata6
Gata4
Gata5
Lpp
Nfe2l3
Nfe2l2
Tmf1
Gli1
Tbx2
En1
En2
Hnf1a
Tbx4
Zfp423
Elf1
Foxb1
Elf2

Name of Genes being analyzed
Elf4
Mxd1
Wt1
Rfx4
Bhlhe41
Sox13
Taf4b
Rfx2
Sox17
Ahr
Sall4
Med14
Smyd1
Sall3
Arid2
Zfp64
Pgr
Trps1
Hoxa1
Bach2
Bach1
Notch3
Pknox1
Pknox2
Sin3a
Etv3
Smad9
Smad5
Alx1
Egf
Mn1
Nkx3-1
Rbak
Gabpa
Nfkbiz
Zscan21
Trp73
E2f7
Esrrg
Rbpjl
Nfatc4
Nr5a1

Name of Genes being analyzed
Neurod4
Esrrb
Tbx21
Rorc
Mitf
Pax7
Pax6
Pax1
Pax3
Pax2
Pax9
Zkscan17
Gfi1
Mzf1
Runx3
Smarca4
Foxd4
Foxd3
Creb1
Srebf1
Sox11
Gmeb2
Irx4
Pou3f2
Ikzf1
Tcf23
Mtf2
Npas3
Nfatc3
Nfil3
Phox2b
Plag1
E2f2
Ddx3x
Taf2
Pou4f1
Trim33
Tsc2
Lmx1a
Nr2f2
Eomes
Wwtr1

Name of Genes being analyzed
Foxo1
Ar
Zfp354a
Elk4
Foxo4
Sall1
Mycn
Maml3
Foxp3
Atm
Uaca
Tbr1
Pml
Lhx3
Atr
Zbtb33
Ptch1
Lhx5
Barhl1
Irx5
Tfap2b
Tfap2e
Rxra
Rxrb
Gli2
Gli3
Zic4
Zic5
Zic2
Zic3
Satb1
Onecut2
Foxn4
Mnat1
Foxn1
Dlx2
Vezf1
sncg
sst
th
vip
xdh

Name of Genes being analyzed
slc17a8
slc5a7
slc6a3
slc6a8
smad3
opalin
pdgfra
palvb
reln
slc17a7
lyve
mfge8
mog
myl14
ndnf
ctss
foxj1
gad1
htr3a
igtp
acta2
alldh1l1
camk2
chat
cldn5
ngef
tiam1
slc1a2
gja1
fbll1

[00442] *seqFISH provides a generalized method to multiplex mRNA imaging in tissues*

[00443] seqFISH with amplification and error correction provides a highly quantitative method to profile hundreds of mRNA species directly in single cells within their native anatomical context. Our method of stripping the probes from the RNA has many advantages. DNAse digestion of probes allows false positives to be rejected as nonspecifically bound probes do not colocalize between different rounds of hybridization (Figure 33A). In addition, the same region of the transcript can be hybridized in every round, allowing seqFISH to efficiently target

mRNAs shorter than 1kb, enabling targeting of most genes. Lastly, seqFISH allows exponential scaling of barcode numbers, thus 4-5 rounds of hybridization can code for hundreds of transcripts with a simple error correction scheme. Theoretically, the entire transcriptome can be coded for with error correction by using 8-9 rounds of hybridization with seqFISH. These advantages of HCR seqFISH allows robust multiplexed RNA detection in tissues, shown here in the mouse brain.

[00444] Ultimately, the multiplexing capability of seqFISH is limited by the amount of optical space within a cell, and not by the coding capacity of the method (supplementary text). We showed previously that super-resolution microscopy can significantly increase the optical space available in the cell for transcription profile imaging, but super-resolution microscopy experiments proved difficult to image in samples thicker than 1μm, and were experimentally cumbersome and time consuming to image (Lubeck and Cai, 2012). A recent development in expansion microscopy as well as correlation methods (Coskun et al., 2016) however offers promise for multiplexing to levels of high transcript density (Chen et al., 2015a; Treweek et al., 2015, Chen et al., 2016). In addition, by labeling subcellular components (i.e., dendrites and axons) with antibodies, the local transcriptome in compartments of the cell can be measured.

[00445] It was observed that, because expression patterns amongst genes are highly correlated, the distinction between large classes of cells can be determined from 10-20 genes, while a finer classification of cell clusters depends on the quantitative measurement of the combinatorial expression patterns of many genes (Figures 34E and F). This correlation amongst genes can be used to "stitch" our seqFISH data with single cell RNAseq data, similar to the approach explored with single cell RNAseq and ISH in Satija et al (Satija et al., 2015) . By correlating seqFISH data to single cell RNA-seq expression data, cells types identified based on RNA-seq can be "mapped" back into our seqFISH data.

[00446] As shown here, seqFISH with hundreds of genes in tissues can become a general and widely used tool to answer a wide range of fundamental questions in biology and medicine. For neuroscience, by combining the insights into the spatial organization of transcription provided by seqFISH with connectomics and electrophysiological measurements, we can obtain a comprehensive understanding of the molecular basis of the neuroanatomy of the brain.

EXAMPLE 5

Supplementary Experimental Procedure for Brain Slide Analysis

[00447] **Probe Design.** Genes were selected from the Allen Brain Atlas database. We identified genes that are heterogeneously expressed in coronal sections containing the hippocampus at Bregma coordinates -2.68 mm anterior. Using the ABA region definitions, we break down the voxels representing the ABA data in those brain sections into 160 distinct regions and average the expression values within each region. We selected 100 genes that had high variances across these distinct regions and that also had low-medium expression levels. These genes included transcription factors and signaling pathways components as well as ion channels and other functional genes. Lastly, we chose 25 genes from single cell RNA-seq data that were enriched in certain cell types. Briefly, the design criteria used were 1) constant regions of all spliced isoforms were identified, 2) Masked regions of UCSC genome were removed from possible probe design, 3) 35mer sequences were tiled 4nt apart, 4) sets of non-overlapping probes with tightest GC range around 55% were found, 5) probes were blasted for off-target hits. Any probe with an expected total off-target copy number of more than 5000 was dropped. Once all possible probes for every target gene was acquired, the probe set oligo-pool was optimized using the following criteria: 1) Expected # of off-target hits for entire probe pool was calculated, 2) probes were sequentially dropped from genes until any off-target gene was hit by no more than 6 probes from entire pool, 3) HCR adapters were added to designed probes and 10nt in either direction of the adapter junction was blasted and screened for off-target hits, 4) probe pools were searched for regions of 18mer complementary, 5) the probe sets for a given transcript was refined down to 24 probes by dropping probes in order of the expected number of off-target hits, 6) Cutting sites and hybridization specific primers were added to probes.

[00448] **Probe Generation.** All oligoarray pools were purchased as 92k synthesis from Customarray Inc. Probes were amplified from array-synthesized oligo pool), with the following modifications: (i) a 35nt RNA-targeting sequence for in situ hybridization, (ii) a 35nt HCR initiator sequence designed to initiate one color of 5 possible HCR polymers, (iii) two hybridization specific flanking primer sequences to allow PCR amplification of the probe set and (iv) EcoRI (5'-GAATTC—3') and KpnI (5'-GGTACC-3') sites for cutting out flanking primers to reduce probe size. Ethanol precipitation was used to purify the final digested probes.

[00449] **Brain extraction and sample mounting.** C57BL/6 with Ai6 Cre-reporter (uncrossed) (Jackson Labs, SN: 007906) female mice aged 50-80 days were anesthetized with isoflurane according to institute protocols (protocol #1701-14). No randomization of mice was used and blinding was not necessary as the study was exploratory. Mice were perfused for 8 minutes with perfusion buffer (10U/ml heparin, 0.5% $NaNO_2$ (w/v) in 0.1M PBS at 4°C). Mice were then perfused with fresh 4% PFA\0.1M PBS buffer at 4C for 8 minutes. The mouse brain was dissected out of the skull and immediately placed in a 4% PFA buffer for 2 hours at room temperature under gentle mixing. The brain was then immersed in 4C 30% RNAse-free Sucrose (Amresco 0335-2.5KG)\1x PBS until the brain sank. After the brain sank, the brain was frozen in an dry ice\isopropanol bath in OCT media and stored at -80C. Fifteen micron sections were cut using a cryotome and immediately placed on an aminosilane modified coverslip.

[00450] **Sample permeabilization, hybridization, and Imaging.** Brain sections mounted to coverslips were permeabilized in 4C 70% EtOH for 12-18 hours. Brains were further permeabilized by the addition of rnase-free 8% SDS (Ambion AM9822) for 10 minutes. Samples were rinsed to remove SDS, desiccated and a hybridization chamber (Grace Bio-Labs 621505) was adhered around the brain section. Samples were hybridized overnight at 37C with Split Color PGK1 Probes in Hybridization Buffer (2X SSC (Invitrogen 15557-036), 10% Formaldehyde (v/v) (Ambion AM9344), 10% Dextran Sulfate (Sigma D8906), 2mM Vanadyl Ribonucleoside Complex (VRC; NEB S1402S) in Ultrapure water (Invitrogen 10977-015)). Samples were washed in 30% Wash Buffer (WBT: 2X SSC, 30% Formaldehyde (v/v)] 10% Dextran Sulfate, 0.1% Triton-X 100 (Sigma X-100), 2mM VRC in Ultrapure water) for 30 minutes. While washing aliquoted HCR hairpins (Molecular Instruments Inc) were heated to 95C for 1.5 minutes and allowed to cool to RT for 30 minutes. HCR hairpins were diluted to a concentration of 120nM per hairpin in amplification buffer (2X SSC, 10% Dextran Sulfate) and added to washed tissue for 45 minutes. Following amplification, samples were washed in the same 30% WBT for at least 10 minutes to remove excess hairpins. Samples were stained with DAPI and submerged in pyranose oxidase antibleaching buffer. Sample port covers were closed with a glass coverslip or a transparent polycarbonate sheet to exclude oxygen.

[00451] Samples were imaged using a standard epifluorescence microscope (Nikon Ti Eclipse with custom built laser assembly) for the 125-gene experiment. Exposures times were

200 ms for cy7 and alexa 488 channels and 100 ms for alexa 647, alexa 594, and cy3b channels. For the 249-gene experiment, a Yokogawa CSU-W1 spinning disk confocal unit attached to an Olympus IX-81 base was used for imaging. The exposure times were 500 ms for each channel. At this stage, intact and accessible mRNA should always appear in two channels. If the RNA was deemed to be intact, DAPI data was collected in this hybridization. Samples were digested with DNAse I (Roche 04716728001) for 4 hours at room temperature on the scope. Following DNAse I the sample was washed several times with 30% WBT and hybridized overnight with 70% Formamide HB and the experiment probes at 1 nM concentration per probe sequence at room temperature. Samples were again washed and amplified as before. Barcode digits were developed by repeating this cycle with the appropriate probes for each hybridization. Fluorescent Nissl stain (ThermoFisher N-21480) was collected at the end of the experiment along with images of multispectral beads to aid chromatic aberration corrections.

[00452] **Image Processing.** To remove the effects of chromatic aberration, the multispectral beads were first used to create geometric transforms to align all fluorescence channels. Next, the background illumination profile of every fluorescence channel was mapped using a morphological image opening with a large structuring element. These illumination profile maps were used to flatten the illumination in post-processing resulting in relatively uniform background intensity and preservation of the intensity profile of fluorescent points. The background signal was then subtracted using the imagej rolling ball background subtraction algorithm with a radius of 3 pixels. Finally, the calculated geometric transforms were applied to each channel respectively. The 150 pixel border region around the image was ignored in all analysis to avoid errors from edge effects of illumination.

[00453] **Image Registration.** The processed images were then registered by first taking a maximum intensity projection along the z direction in each channel. All of the maximum projections of the channels of a single hybridization were then collapsed resulting in 4 composite images containing all the points in a particular round of hybridization. Each of these composite images of hybridization 1-3 were then cross-correlated individually with the composite image of hybridization 4 and the position of the maxima of the cross-correlation was used as the translation factor to align hybridizations 1-3 to hybridization 4.

[00454] **Cell Segmentation.** For cells in the cortex, the cells were segmented manually using the DAPI images taken in the first round of hybridization and the fluorescent nissl stain taken at the end of the experiment. Furthermore, the density of the point cloud surrounding a cell was taken into account when forming cell boundaries, especially in cells that did not stain with the nissl stain. For the hippocampus, the cells were segmented by first manually selecting the centroid in 3D of each DAPI signal of every cell. Transcripts were first assigned based on nearest centroids. These point clouds were then used to refine the centroid estimate and create a 3D voronoi tessellation with a 10% boundary-shrinking factor to eliminate ambiguous mRNA assignments from neighboring cells.

[00455] **Barcode calling.** The potential mRNA signals were then found by LOG filtering the registered images and finding points of local maxima above a specified threshold value. Once all potential points in all channels of all hybridizations were obtained, dots were matched to potential barcode partners in all other channels of all other hybridizations using a 1 pixel search radius to find symmetric nearest neighbors. Point combinations that constructed only a single barcode were immediately matched to the on-target barcode set. For points that matched to construct multiple barcodes, first the point sets were filtered by calculating the residual spatial distance of each potential barcode point set and only the point sets giving the minimum residuals were used to match to a barcode. If multiple barcodes were still possible, the point was matched to its closest on-target barcode with a hamming distance of 1. If multiple on target barcodes were still possible, then the point was dropped from the analysis as an ambiguous barcode. This procedure was repeated using each hybridization as a seed for barcode finding and only barcodes that were called similarly in at least 3 out of 4 rounds were used in the analysis. The number of each barcode was then counted in each of the assigned cell volumes and transcript numbers were assigned based on the number of on-target barcodes present in the cell volume. All image processing and image analysis code can be obtained upon request.

[00456] **Clustering.** To cluster the dataset with 14,908 cells and 125 genes profiled, we first z-score normalized the data based on gene expression. Once the single cell gene expression data is converted into z-scores, we compute a matrix of cell-to-cell correlations using Pearson correlation coefficients. Then hierarchical clustering with Ward linkage is performed on the cell-to-cell correlation data with cells in the center field of view. The cluster definitions are then

propagated to the remaining cells using a random forest machine learning algorithm. To analyze the robustness of individual clusters, a random forest model was trained using varying subsets of the data and used to predict the cluster assignment of the remaining cells. A bootstrap analysis by dropping different sets of cells was performed in increments (Figure 42). To determine the effect of dropping out genes on the accuracy of the clustering analysis, we used a random forest decision tree to learn the cluster definition based on the 125 gene data. Then we ask the decision tree to re-compute the cluster assignment on cell-to-cell correlation matrices with fewer and fewer genes (Figure 34F, green line). Bootstrap resampling was also performed with this analysis (Figure 34F, bluelines). The PCA and tSNE analysis were performed using the same cell-to-cell z-scored Pearson correlation matrix. The cell-to-cell correlation in Figure 34E was calculated with increasing number of principal components dropped (have their eigenvalues set to zero). The cluster assignment accuracy is again computed through the random forest decision tree.

[00457] Optical Space for Barcodes in Cells. The theoretical upper limit for the number of barcodes that can be identified accurately within a cells primarily depends on the volume of the cell. As mRNA spots are diffraction limited, if a microscope is configured to have sub-diffraction limited pixel size, the ability to identify smFISH signal without any super-resolution would require no two mRNA signals to be immediately adjacent to each other in x, y or z dimension. These minimum required voxels are called "coding voxels." The absolute upper limit of the number of transcripts that can be coded unambiguously without any super-resolution methods is solely a function of the number of coding voxels present in a cell. Assuming a diffraction limit of λ um and a resolution of z um in the z direction, there exists $\frac{V}{(3\lambda)^2 z}$ coding voxels per cell, where V is the volume of the cell in microns. In the seqFISH method, we use 5 or more channels to hold mRNA spots which would increase the total number of coding voxels by a multiplicative factor equal to the number of channels used for barcoding. Therefore,

$$\#B = \frac{F\,V}{(3\lambda)^2 z}$$

where #B is the maximum number of unambiguous barcodes a cell can hold, and F is the number of channels used. As mammalian cells range from about 500 – 4000 microns in volume, these cells can accommodate roughly between 6100 – 49,000 barcodes assuming 5 fluorescence

channels are being used, the diffraction limit is 0.3 um, and the z resolution is 0.5 um. In principle, this calculation would provide the total number of perfectly discernible spots a cell can accommodate. In our actual experimental data, we have some amount of dropped barcodes due to ambiguity in barcode assignment due to spot overlaps. This is one of the main factors that reduces the efficiency of seqFISH as compared to single transcript detection (i.e., smFISH or smHCR). Expansion microscopy could further increase the number of coding voxels in a cell by the expansion factor leading to fewer drops and imaging of denser transcripts.

[00458] Additional background information can be found in the following references, each of which is hereby incorporated by reference in its entirety.

[00459] Beliveau, B.J., Joyce, E.F., Apostolopoulos, N., Yilmaz, F., Fonseka, C.Y., McCole, R.B., Chang, Y., Li, J.B., Senaratne, T.N., Williams, B.R., et al. (2012). Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. Proc. Natl. Acad. Sci. U. S. A. 109, 21301–21306.

[00460] Betzig, E., Patterson, G.H., Sougrat, R., Lindwasser, O.W., Olenych, S., Bonifacino, J.S., Davidson, M.W., Lippincott-Schwartz, J., and Hess, H.F. (2006). Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. Science 313, 1642–1645.

[00461] Breiman, L. (2001). Random Forests. Mach. Learn. 45, 5–32.

[00462] Bose S, Wan Z, Carr A, Rizvi AH, Vieira G, Pe'er D, Sims PA. Scalable microfluidics for single-cell RNA printing and sequencing. Genome Biol. 2015 Jun 6;16:120.

[00463] Buenrostro JD, Araya CL, Chircus LM, Layton CJ, Chang HY, Snyder MP, Greenleaf WJ. Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. Nat Biotechnol. 2014 Jun;32(6):562-8.

[00464] Cajigas, I.J., Tushev, G., Will, T.J., Dieck, S. tom, Fuerst, N., and Schuman, E.M. (2012). The Local Transcriptome in the Synaptic Neuropil Revealed by Deep Sequencing and High-Resolution Imaging. Neuron 74, 453–466.

[00465] Junyue Cao, Jonathan S. Packer, Vijay Ramani, Darren A. Cusanovich, Chau Huynh, Riza Daza, Xiaojie Qiu, Choli Lee, Scott N. Furlan, Frank J. Steemers, Andrew Adey, Robert H. Waterston, Cole Trapnell, Jay Shendure. Comprehensive single cell transcriptional

profiling of a multicellular organism by combinatorial indexing. bioRxiv 104844; doi: https://doi.org/10.1101/104844

[00466] Cembrowski, M.S., Bachman, J.L., Wang, L., Sugino, K., Shields, B.C., and Spruston, N. (2016). Spatial Gene-Expression Gradients Underlie Prominent Heterogeneity of CA1 Pyramidal Neurons. Neuron 89, 351–368.

[00467] Cenquizca, L.A., and Swanson, L.W. (2007). Spatial organization of direct hippocampal field CA1 axonal projections to the rest of the cerebral cortex. Brain Res. Rev. 56, 1–26.

[00468] Chen, F., Tillberg, P.W., and Boyden, E.S. (2015a). Expansion microscopy. Science 347, 543–548.

[00469] Chen, F., Wassie, A.T., Cote, A.J., Sinha, A., Alon, S., Asano, S., Daugharthy, E.R., Chang, J.-B., Marblestone, A., Church, G.M., Raj, A., Boyden, E.S., 2016. Nanoscale imaging of RNA with expansion microscopy. Nat Meth advance online publication.

[00470] Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015b). Spatially resolved, highly multiplexed RNA profiling in single cells. Science 348, aaa6090.

[00471] Choi, H.M.T., Beck, V.A., and Pierce, N.A. (2014). Next-Generation in Situ Hybridization Chain Reaction: Higher Gain, Lower Cost, Greater Durability. ACS Nano 8, 4284–4294.

[00472] Darmanis, S., Sloan, S.A., Zhang, Y., Enge, M., Caneda, C., Shuer, L.M., Gephart, M.G.H., Barres, B.A., and Quake, S.R. (2015). A survey of human brain transcriptome diversity at the single cell level. Proc. Natl. Acad. Sci. 112, 7285–7290.

[00473] Dong, H.-W., Swanson, L.W., Chen, L., Fanselow, M.S., and Toga, A.W. (2009). Genomic–anatomic evidence for distinct functional domains in hippocampal field CA1. Proc. Natl. Acad. Sci. 106, 11794–11799.

[00474] Donner Y, Feng T, Benoist C, Koller D. Imputing gene expression from selectively reduced probe sets. Nat Methods. 2012 Nov;9(11):1120-5.

[00475] Duan et al. L1000CDS2: LINCS L1000 characteristic direction signatures search engine. npj Systems Biology and Applications (2016) 2, 16015;

[00476] Fan, Y., Braut, S.A., Lin, Q., Singer, R.H., and Skoultchi, A.I. (2001). Determination of transgenic loci by expression FISH. Genomics 71, 66–69.

[00477] Fanselow, M.S., and Dong, H.-W. (2010). Are the dorsal and ventral hippocampus functionally distinct structures? Neuron 65, 7–19.

[00478] Femino, A.M., Fay, F.S., Fogarty, K., and Singer, R.H. (1998). Visualization of Single RNA Transcripts in Situ. Science 280, 585–590.

[00479] Fulton DL, Sundararajan S, Badis G, Hughes TR, Wasserman WW, Roach JC, Sladek R. TFCat: the curated catalog of mouse and human transcription factors. Genome Biol. 2009;10(3):R29. doi: 10.1186/gb-2009-10-3-r29.

[00480] Habib, N., Li, Y., Heidenreich, M., Swiech, L., Trombetta, J.J., Zhang, F., Regev, A., 2016. Div-Seq: A single nucleus RNA-Seq method reveals dynamics of rare adult newborn neurons in the CNS. bioRxiv 045989.

[00481] Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science. 2009 Apr 10;324(5924):218-23.

[00482] Jung, M.W., Wiener, S.I., and McNaughton, B.L. (1994). Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. J. Neurosci. 14, 7347–7356.

[00483] Ke, R., Mignardi, M., Pacureanu, A., Svedlund, J., Botling, J., Wählby, C., and Nilsson, M. (2013). In situ sequencing for RNA analysis in preserved tissue and cells. Nat. Methods 10, 857–860.

[00484] Kishi, T., Tsumori, T., Yokota, S., and Yasui, Y. (2006). Topographical projection from the hippocampal formation to the amygdala: A combined anterograde and retrograde tracing study in the rat. J. Comp. Neurol. 496, 349–368.

[00485] Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. Cell 161, 1187–1201.

[00486] Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Yang, J.L., Ferrante, T.C., Terry, R., Jeanty, S.S.F., Li, C., Amamoto, R., et al. (2014). Highly Multiplexed Subcellular RNA Sequencing in Situ. Science 343, 1360–1363.

[00487] Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. Nature 445, 168–176.

[00488] Levesque MJ, Ginart P, Wei Y, Raj A. Visualizing SNVs to quantify

[00489] allele-specific expression in single cells. Nat Methods. 2013 Sep;10(9):865-867.

[00490] Lubeck, E., and Cai, L. (2012). Single-cell systems biology by super-resolution imaging and combinatorial labeling. Nat. Methods 9, 743–748.

[00491] Lubeck, E., Coskun, A.F., Zhiyentayev, T., Ahmad, M., and Cai, L. (2014). Single-cell in situ RNA profiling by sequential hybridization. Nat. Methods 11, 360–361.

[00492] Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell 161, 1202–1214.

[00493] Madisen, L., Zwingman, T.A., Sunkin, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L., Palmiter, R.D., Hawrylycz, M.J., Jones, A.R., et al. (2010). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. Nat. Neurosci. 13, 133–140.

[00494] Madisen, L., Mao, T., Koch, H., Zhuo, J., Berenyi, A., Fujisawa, S., Hsu, Y.-W.A., Iii, A.J.G., Gu, X., Zanella, S., et al. (2012). A toolbox of Cre-dependent optogenetic transgenic mice for light-induced activation and silencing. Nat. Neurosci. 15, 793–802.

[00495] Mellis IA, Gupte R, Raj A, Rouhanifard SH. Visualizing adenosine-to-inosine RNA editing in single mammalian cells. Nat Methods. 2017 Jun 12. doi: 10.1038/nmeth.4332.

[00496] Miller, JA. Jason Nathanson, Daniel Franjic, Sungbo Shim, Rachel A. Dalley, Sheila Shapouri, Kimberly A. Smith, Susan M. Sunkin, Amy Bernard, Jeffrey L. Bennett, Chang-Kyu Lee, Michael J. Hawrylycz, Allan R. Jones, David G. Amaral, Nenad Sestan, Fred H. Gage, Ed S. Lein (2013). Conserved molecular signatures of neurogenesis in the hippocampal subgranular zone of rodents and primates. Development. 140(22): 4633–4644.

[00497] Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008 Jul;5(7):621-628.

[00498] Muller, R., Stead, M., and Pach, J. (1996). The hippocampus as a cognitive graph. J. Gen. Physiol. 107, 663–694.

[00499] Nagalakshmi, U. et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320, 1344–1349 (2008).

[00500] O'Keefe, J., and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res. 34, 171–175.

[00501] Petrovich, G.D., Canteras, N.S., and Swanson, L.W. (2001). Combinatorial amygdalar inputs to hippocampal domains and hypothalamic behavior systems. Brain Res. Brain Res. Rev. 38, 247–289.

[00502] Pitkänen, A., Pikkarainen, M., Nurminen, N., and Ylinen, A. (2000). Reciprocal Connections between the Amygdala and the Hippocampal Formation, Perirhinal Cortex, and Postrhinal Cortex in Rat: A Review. Ann. N. Y. Acad. Sci. 911, 369–391.

[00503] Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA Synthesis in Mammalian Cells. PLoS Biol 4, e309.

[00504] Risold, P.Y., and Swanson, L.W. (1996). Structural evidence for functional domains in the rat hippocampus. Science 272, 1484–1486.

[00505] Alexander B Rosenberg, Charles Roco, Richard A Muscat, Anna Kuchina, Sumit Mukherjee, Wei Chen, David J Peeler, Zizhen Yao, Bosiljka Tasic, Drew L Sellers, Suzie H Pun, Georg Seelig. Scaling single cell transcriptomics through split pool barcoding bioRxiv 105163; doi: https://doi.org/10.1101/105163

[00506] Rust, M.J., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nat Meth 3, 793–796.

[00507] Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., Regev, A., 2015. Spatial reconstruction of single-cell gene expression data. Nat Biotech 33, 495–502.

[00508] Saunders, R.C., Rosene, D.L., and Van Hoesen, G.W. (1988). Comparison of the efferents of the amygdala and the hippocampal formation in the rhesus monkey: II. Reciprocal and non-reciprocal connections. J. Comp. Neurol. 271, 185–207.

[00509] S. Shah, E. Lubeck, W. Zhou, L. Cai, In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. Neuron 92, 342-357 (2016).

[00510] Singer ZS, Yong J, Tischler J, Hackett JA, Altinok A, Surani MA, Cai L, Elowitz MB. Dynamic heterogeneity and DNA methylation in embryonic stem cells. Mol Cell. 2014 Jul 17;55(2):319-31.

[00511] Shah, S., Lubeck, E., Schwarzkopf, M., He, T., Greenbaum, A., Sohn, C. ho, Lignell, A., Choi, H.M.T., Gradinaru, V., Pierce, N.A., Cai, L., 2016. Single-molecule RNA detection at depth via hybridization chain reaction and tissue hydrogel embedding and clearing. Development dev.138560. doi:10.1242/dev.138560

[00512] Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., Mollbrink, A., Linnarsson, S., Codeluppi, S., Borg, Å., Pontén, F., Costea, P.I., Sahlén, P., Mulder, J., Bergmann, O., Lundeberg, J., Frisén, J., 2016. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science 353, 78–82. doi:10.1126/science.aaf2403

[00513] Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T., et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat. Neurosci. advance online publication.

[00514] Thompson, C.L., Pathak, S.D., Jeromin, A., Ng, L.L., MacPherson, C.R., Mortrud, M.T., Cusick, A., Riley, Z.L., Sunkin, S.M., Bernard, A., et al. (2008). Genomic Anatomy of the Hippocampus. Neuron 60, 1010–1021.

[00515] Treweek, J.B., Chan, K.Y., Flytzanis, N.C., Yang, B., Deverman, B.E., Greenbaum, A., Lignell, A., Xiao, C., Cai, L., Ladinsky, M.S., et al. (2015). Whole-body tissue stabilization and selective extractions via tissue-hydrogel hybrids for high-resolution intact circuit mapping and phenotyping. Nat. Protoc. 10, 1860–1896.

[00516] Van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. J. Mach. Learn. Res. 9, 85.

[00517] Witter, M.P. (1993). Organization of the entorhinal—hippocampal system: A review of current anatomical data. Hippocampus 3, 28–44.

[00518] Witter, M.P., and Amaral, D.G. (1991). Entorhinal cortex of the monkey: V. Projections to the dentate gyrus, hippocampus, and subicular complex. J. Comp. Neurol. 307, 437–459.

[00519] Yang, B., Treweek, J.B., Kulkarni, R.P., Deverman, B.E., Chen, C.-K., Lubeck, E., Shah, S., Cai, L., and Gradinaru, V. (2014). Single-Cell Phenotyping within Transparent Intact Tissue through Whole-Body Clearing. Cell.

[00520] Yang SM, Alvarez DD, Schinder AF. (2015). Reliable Genetic Labeling of Adult-Born Dentate Granule Cells Using Ascl1 CreERT2 and Glast CreERT2 Murine Lines. J Neurosci. 35(46):15379-90.

[00521] Yi, F., Catudio-Garrett, E., Gábriel, R., Wilhelm, M., Erdelyi, F., Szabo, G., Deisseroth, K., and Lawrence, J. (2015). Hippocampal "cholinergic interneurons" visualized with the choline acetyltransferase promoter: anatomical distribution, intrinsic membrane properties, neurochemical characteristics, and capacity for cholinergic modulation. Front. Synaptic Neurosci. 7.

[00522] Zeisel, A., Manchado, A.B.M., Codeluppi, S., Lönnerberg, P., Manno, G.L., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science aaa1934.

EQUIVALENTS

[00523] Having described some illustrative embodiments of the invention, it should be apparent to those skilled in the art that the foregoing is merely illustrative and not limiting, having been presented by way of example only. Numerous modifications and other illustrative embodiments are within the scope of one of ordinary skill in the art and are contemplated as falling within the scope of the invention. In particular, although many of the examples presented herein involve specific combinations of method acts or system elements, it should be understood that those acts and those elements may be combined in other ways to accomplish the same objectives. Acts, elements, and features discussed only in connection with one embodiment are not intended to be excluded from a similar role in other embodiments. Further, for the one or more means-plus-function limitations recited in the following claims, the means are not intended to be limited to the means disclosed herein for performing the recited function, but are intended to cover in scope any means, known now or later developed, for performing the recited function.

[00524] Use of ordinal terms such as "first", "second", "third", *etc.*, in the claims to modify a claim element does not by itself connote any priority, precedence, or order of one claim

element over another or the temporal order in which acts of a method are performed, but are used merely as labels to distinguish one claim element having a certain name from another element having a same name (but for use of the ordinal term) to distinguish the claim elements. Similarly, use of a), b), *etc.*, or i), ii), *etc.* does not by itself connote any priority, precedence, or order of steps in the claims. Similarly, the use of these terms in the specification does not by itself connote any required priority, precedence, or order.

[00525] The foregoing written specification is considered to be sufficient to enable one skilled in the art to practice the invention. The present invention is not to be limited in scope by examples provided, since the examples are intended as a single illustration of one aspect of the invention and other functionally equivalent embodiments are within the scope of the invention. Various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and fall within the scope of the appended claims. The advantages and objects of the invention are not necessarily encompassed by each embodiment of the invention.

WE CLAIM:

1. A method of barcoding molecular targets, comprising:

identifying N molecular targets in a biological sample, wherein the N molecular targets are immobilized;

associating a unique barcode to each molecular target via n sequential barcoding rounds (where $n \geq 2$), wherein each barcoding round comprises m serial hybridizations of probes collectively bound to the N molecular targets (where $m \geq 2$), wherein each serial hybridization comprises:

contacting one or more groups of probes to a subset of the N molecular targets, the total number of groups of probes corresponding to the number of molecular targets in the subset, wherein probes in each group comprise one or more binding sequences specifically targeting a molecular target in the subset, wherein each probe is capable of generating at least one detectable visual signal representing binding of the probe to a molecular target in the subset, and wherein probes in the one or more groups generate one or more different detectable visual signals corresponding to the number of molecular targets in the subset;

detecting the detectable visual signals that reflect the binding between the one or more groups of probes and the subset of the N molecular targets; and

removing the visual signals, when applicable, prior to the next serial hybridization;

wherein the unique barcode to each molecular target consists of n components, each component is assigned from S unique symbols, where S is an integer that equal to or greater than $\sqrt[n]{N}$; and

optionally removing probes between two barcoding rounds.

2. The method of claim 1, wherein the detecting the detectable visual signals comprises:

capturing, for each serial hybridization round, an image of the detectable visual signals that reflect the binding between the one or more groups of probes and the subset of the N molecular targets.

3. The method of claim 1 or claim 2, further comprising:

generating, for each barcoding round, a composite image by superimposing m images corresponding to the m serial hybridizations, wherein the m images are aligned based on one or more alignment references whose positions remain constant relative to the biological sample.

4. The method of any one of claims 1-3, further comprising:

applying Gaussian analysis to super-localize the detectable visual signals in an image.

5. The method of any one of claims 1-4, further comprising:

decoding the detectable visual signals in each composite image based on the unique barcodes for the N molecular targets and the S unique symbols.

6. The method of any one of claims 1-5, further comprising:

detecting reference visual signals associated with the one or more alignment reference.

7. The method of any one of claims 1-6, wherein the one or more alignment references comprise one or more selected from the group consisting of an oligonucleotide sequence immobilized on the coverslips and detected by a complementary oligo, a common sequence embedded in all probes, a microscopic object, a metal bead, a gold bead, a polystyrene bead, a PCR handle sequence on a primary binding probe, and combinations thereof.

8. The method of any one of claims 1-7,wherein the n sequential barcoding rounds includes x round for error correction, where x is an integer equal or greater than 1; and wherein assigning unique barcodes for each of N molecular targets requires S unique symbols, where S is an integer equal or greater than $\sqrt[n-x]{N}$.

9. The method of any one of claims 1-8, wherein the biological sample comprises a tissue sample, a cell sample, a cell extract sample, a nucleic acid sample, an RNA transcript sample, a

protein sample, an mRNA sample, DNA molecules, protein molecules, RNA and DNA isoform molecules, single nucleotide polymorphism molecules, or combinations thereof.

10. The method of any one of claims 1-9, further comprising:

determining a secondary molecular target that are associated with the *N* molecular targets by contacting the biological sample with molecules specifically binding to the secondary molecular target.

11. The method of any one of claims 1-10, wherein the secondary molecular target comprises one selected from the group consisting of a RNA binding protein molecule, ribosome, a DNA binding protein molecule, a transcription factor, a chromatin binding protein, a protein binding molecule, a scaffold protein, and combinations thereof.

12. The method of any one of claims 1-11, wherein probes in the one or more groups of probes further comprise:

one or more binding sequences each specifically targeting one or more sites within a molecular target in the biological sample; and

n unique readout sequences associated with the one or more binding sequences, wherein, in each barcoding round, only one unique readout sequence is associated with a detectable visual signal for a particular molecular target.

13. The method of any one of claims 1-12, wherein the one or more binding sequences target multiple different sites within the same molecular target.

14. The method of any one of claims 1-13, wherein the one or more binding sequences target multiple different sites within different molecular targets.

15. The method of any one of claims 1-14, wherein each probe comprises one or more of the *n* unique readout sequences.

16. The method of any one of claims 1-15, wherein at least one of the *n* unique readout sequences is located in an overhang sequence directly connected to the binding sequence of a probe.

17. The method of any one of claims 1-16, wherein at least one of the *n* unique readout sequences is indirectly connected to the binding sequence of a probe via one or more intermediate molecules.

18. The method of any one of claims 1-17, wherein the one or more intermediate molecules comprise an RNA bridge probe, a DNA bridge probe, a protein bridge probe, a probe for hybridization chain reaction (HCR), a hairpin nucleic acid probe, an HCR initiator, an HCR polymer, or combinations thereof.

19. The method of claim any one of claims 1-18, wherein the one or more binding sequences specifically target one or more non-nucleic acid sites in the molecular target, and wherein the *n* unique readout sequences comprising nucleic acid sequences that are directly or indirectly connected to the binding sequences of the probes.

20. The method of any one of claims 1-19, wherein the detectable visual signal is connected to the binding sequence of a probe or an intermediate molecule via a cleavable linker.

21. The method of any one of claims 1-20, wherein the one or more binding sequences comprises a peptide sequence binding to a specific antigen within a particular molecular target, an aptmer, or click chemistry group.

22. The method of any one of claims 1-21, wherein the *S* unique symbols comprise colors, numbers, letters, shapes, or combinations thereof.

23. The method of any one of claims 1-22, wherein, for each serial hybridization, the one or more groups of probes to a non-overlapping subset of the *N* molecular targets.

24. A method of hybridization analysis of binding between labeled probes and molecular targets in a biological sample, comprising:

generating multiple composite images of labeled probes bound to a plurality of molecular targets in the biological sample, wherein each composite image is generated from a plurality of images of labeled probes collectively bound to the plurality of molecular targets, wherein the plurality of molecular targets are immobilized within the biological sample, and wherein each image of the plurality of images reveals:

labeled probes bound to a subset of molecular targets within the plurality of molecular targets, wherein the labeled probes comprise one or more groups of probes, the total number of groups of probes corresponding to the number of molecular targets in the subset, wherein probes in each group comprise one or more binding sequences specifically targeting a molecular target in the subset, and wherein each labeled probe is capable of generating a visual signal representing binding of the probe to a molecular target; and

one or more reference targets whose positions remain constant in the biological sample for aligning the plurality of images.

25. The method of claim 24, wherein the biological sample comprises a tissue sample, a cell sample, a cell extract sample, a nucleic acid sample, an RNA transcript sample, a protein sample, an mRNA sample, DNA molecules, protein molecules, RNA and DNA isoform molecules, single nucleotide polymorphism molecules, or combinations thereof.

26. The method of claim 24 or claim 25, wherein, in each image, the labelled probes bind to a non-overlapping subset of molecular targets within the plurality of molecular targets.

27. The method of any one of claims 24-26, further comprising:

contacting the one or more groups of probes with the subset of molecular targets of the plurality of molecular targets;

detecting visual signals that reflect the binding between the one or more groups of probes and molecular targets in the subset; and

removing the visual signals, when applicable, prior to a next round of hybridization of labeled probes binding to a new subset of molecular targets within the plurality of molecular targets.

28. The method of any one of claims 24-27, further comprising:

detecting reference visual signals associated with the one or more alignment references.

29. The method of any one of claims 24-28, further comprising:

aligning the plurality of images based on the positions of the one or more alignment references.

30. A sequential hybridization method, comprising:

identifying a plurality of target genes; and

associating, via sequential hybridization of binding probes to the plurality of target genes, a first plurality of unique codes with the plurality of target genes, wherein each target gene in the plurality of target genes is represented by a unique code in the first plurality of unique codes, wherein the sequential hybridization comprises n rounds of hybridization (where $n \geq 2$), and wherein each round of hybridization in n rounds of hybridization comprises:

contacting the plurality of target genes with a plurality of binding probes, wherein each probe in the plurality of binding probes comprises:

a binding sequence that specifically binds a target sequence in a gene in the plurality of target genes, wherein target genes from the plurality of target genes are spatially transfixed from each other, and wherein each probe is capable of emitting a detectable visual signal upon binding of the probe to a target sequence;

detecting visual signals that reflect the binding between the plurality of binding probes and the plurality of target genes; and

removing the visual signals, when applicable, prior to the next round of hybridization;

wherein probes used in the n rounds of hybridization are capable of emitting at least F types of detectable visual signals (where $F \geq 2$ and F^n is greater than the number of target genes in the plurality of target genes),

wherein a unique code in the first plurality of unique codes for a target gene consists of n components,

wherein each component is determined by visual signals that reflect the binding between binding probes and the target gene during one of the n rounds of hybridization,

wherein the n rounds of hybridization include m error correction round ($m \geq 1$),

wherein a second plurality of unique codes for the plurality of target genes is generated after the m error correction round is removed from the n rounds of hybridization, and

wherein each unique code in the second plurality of unique codes consists of $(n - m)$ components and uniquely represents a target gene in the plurality of target genes.

31. A hybridization method, comprising:

identifying a plurality of target genes;

performing sequential hybridization of binding probes to the plurality of target genes, wherein the sequential hybridization comprises n rounds of hybridization (where $n \geq 2$), and wherein each round of hybridization in n rounds of hybridization comprises:

contacting the plurality of target genes with a plurality of binding probes, wherein each probe in the plurality of binding probes comprises:

a binding sequence that specifically binds a target sequence in a gene in the plurality of target genes, wherein target genes from the plurality of target genes are spatially transfixed from each other, and wherein each probe is capable of emitting a detectable visual signal upon binding of the probe to a target sequence;

detecting visual signals that reflect the binding between the plurality of binding probes and the plurality of target genes, wherein each target gene in the plurality of target genes is represented by visual signals that are unique for the target gene, and wherein

probes used in the n rounds of hybridization are capable of emitting at least F types of detectable visual signals (where $F \geq 2$, and F^n is greater than the number of target genes in the plurality of target genes); and

removing the visual signals, when applicable, prior to the next round of hybridization; and

performing serial hybridizations against one or more serial target genes, wherein the expression level of each serial target gene is above a predetermined threshold value, wherein each serial hybridization comprises:

contacting the one or more serial target genes with a plurality of binding probes, wherein each probe in the plurality of binding probes comprises:

a binding sequence that specifically binds a target sequence in a serial target gene in the one or more serial target genes, wherein one or more serial target genes are spatially transfixed from each other,

wherein each probe is capable of emitting a detectable visual signal upon binding of the probe to the target sequence, and wherein probes binding to target sequences in the same serial target gene emit the same detectable visual signals; and

detecting visual signals that reflect the binding between the plurality of binding probes and the one or more serial target gene.

32. The sequential hybridization method of claim 30 or claim 31, wherein each component of a n-component unique code in the first plurality of unique codes is assigned a numerical value that corresponds to one of the at least F types of detectable visual signals; and wherein at least one component of the n-component unique code is determined based on the numerical values of all or some of the other n-1 components.

33. The sequential hybridization method of any one of claims 30-32, wherein the n-component unique code is determined as:

$\{j_1, j_2, \ldots (a_1*j_1 + a_2*j_2 \ldots + a_n*j_n + C) \text{mod } F, \ldots, j_n\}$,

wherein j_1 is a numerical value that corresponds the detectable visual signals used in the first round of hybridization, j_2 is a numerical value that corresponds the detectable visual signals used in the second round of hybridization, and j_n is a numerical value that corresponds the detectable visual signals used in the *n*th round of hybridization; and

wherein $j_1, j_2, \ldots j_n, a_1, a_2, \ldots a_n$ and n are none zero integers and C is an integer.

a) spatial barcodes
b) spectral barcodes
c) temporal sequential barcodes
Scales:
$_FP_N$
$_FC_N$
seqFISH
F^N
Max capacity:
(assuming 5 dyes)
720
31
5^7=78,125
5^3=125
Coding scheme
4-5 colors
Probes needed
20-40
20-40
10-20
rehybs probes

FIG. 1

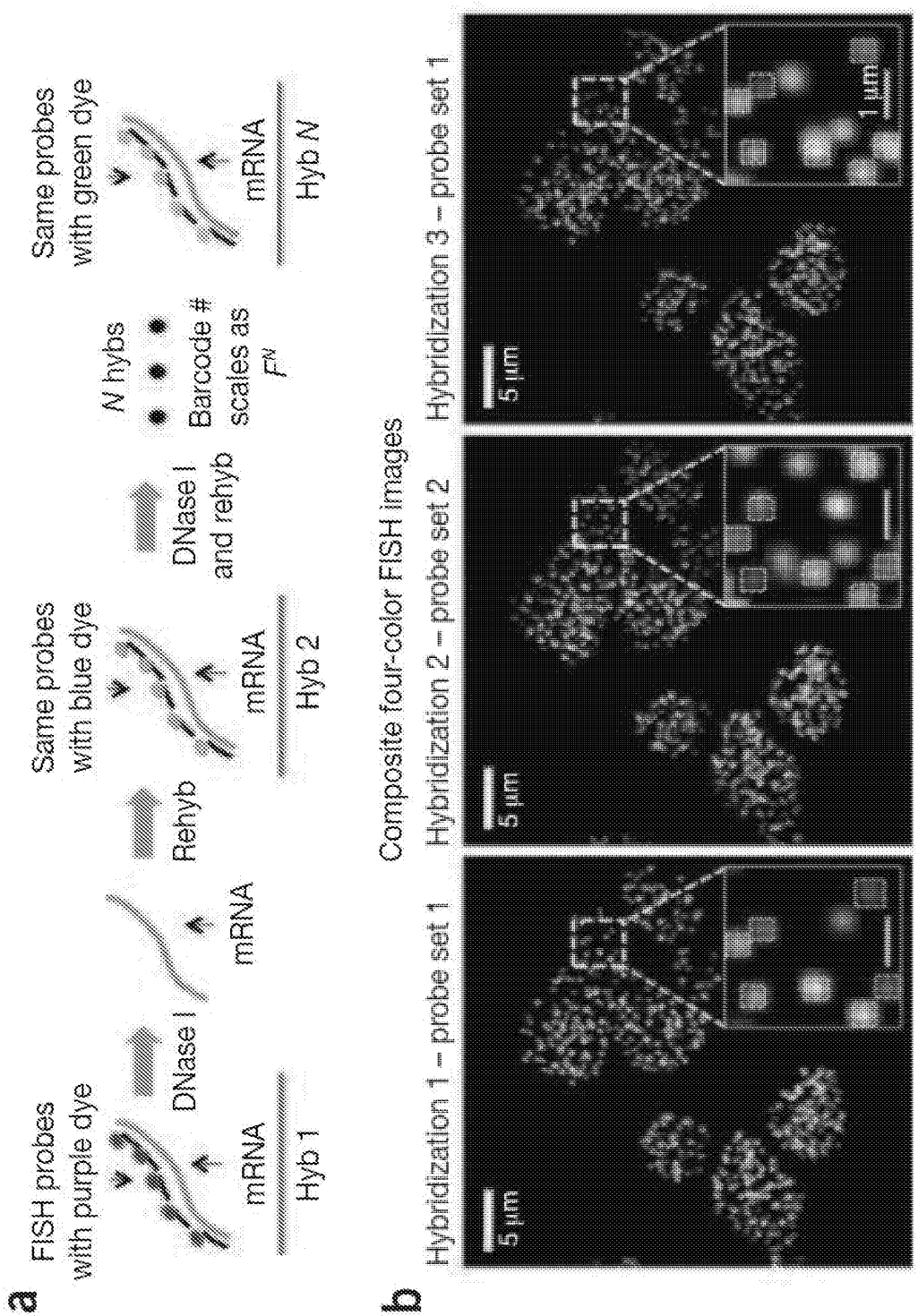

a
FISH probes with purple dye
DNase I
mRNA
Hyb 1
Rehyb
Same probes with blue dye
mRNA
Hyb 2
DNase I and rehyb
N hybs
Barcode # scales as F^N
Same probes with green dye
mRNA
Hyb N
b
Composite four-color FISH images
Hybridization 1 – probe set 1
Hybridization 2 – probe set 2
Hybridization 3 – probe set 1
5 μm
1 μm

FIG. 2

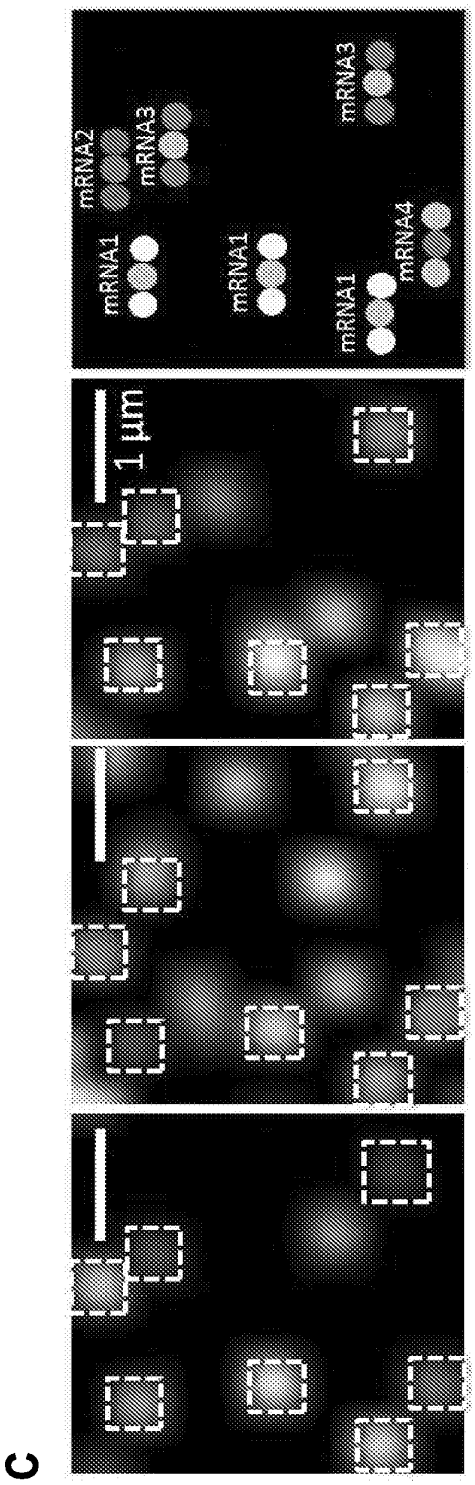


FIG. 2 (Continued)

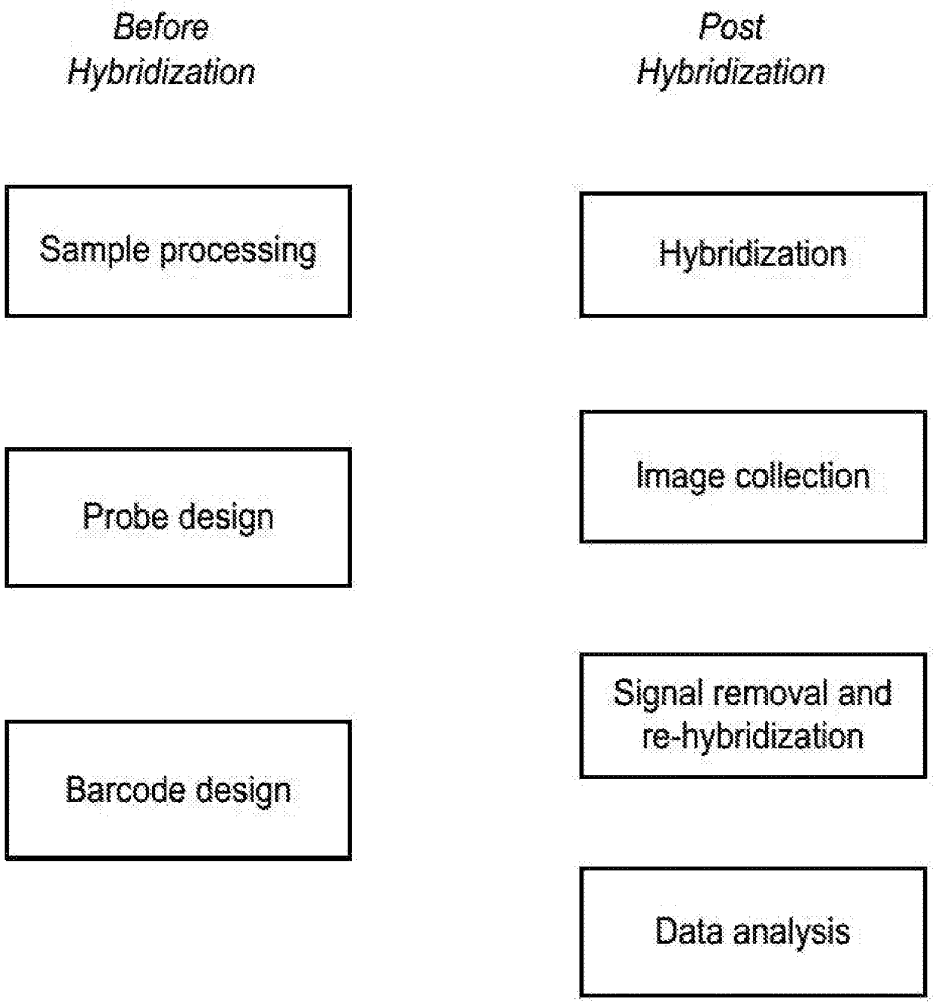

Before Hybridization
Post Hybridization
Sample processing
Probe design
Barcode design
Hybridization
Image collection
Signal removal and re-hybridization
Data analysis

FIG. 3A

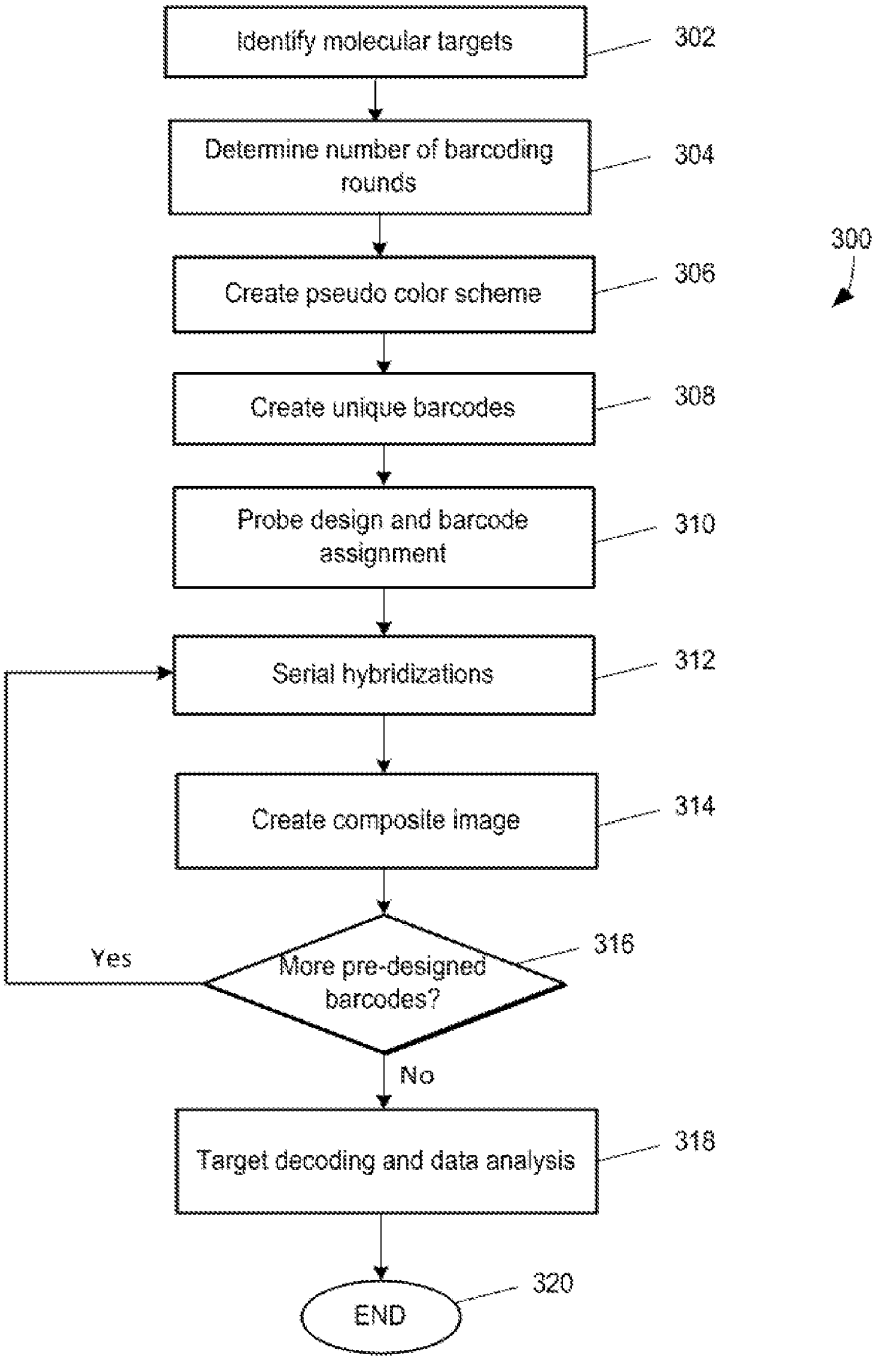

Identify molecular targets
302
Determine number of barcoding rounds
304
300
Create pseudo color scheme
306
Create unique barcodes
308
Probe design and barcode assignment
310
Serial hybridizations
312
Create composite image
314
Yes
More pre-designed barcodes?
316
No
Target decoding and data analysis
318
END
320

FIG. 3B

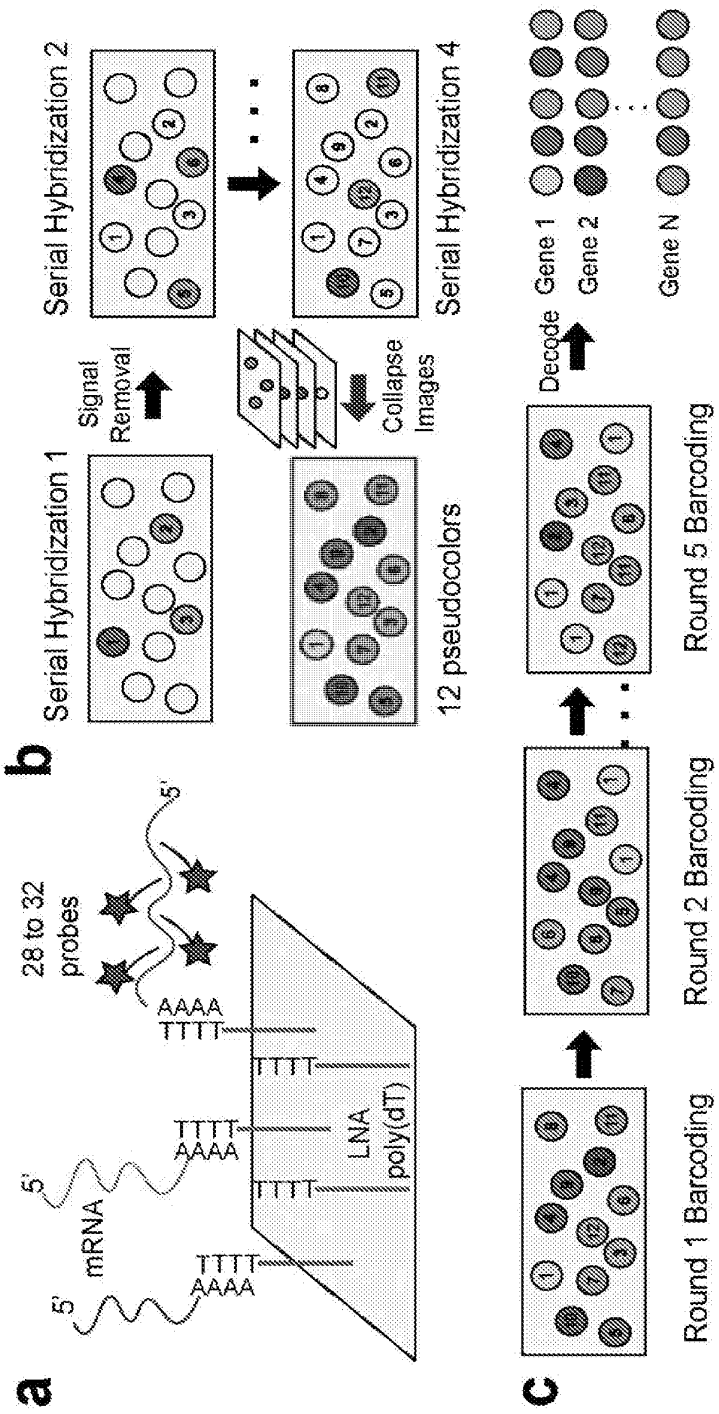

a
5'
mRNA
AAAA
5'
AAAA
5'
AAAA
LNA
poly(dT)
28 to 32
probes
5'
b
Serial Hybridization 1
Signal
Removal
Serial Hybridization 2
Collapse
Images
12 pseudocolors
Serial Hybridization 4
c
Round 1 Barcoding
Round 2 Barcoding
Round 5 Barcoding
Decode
Gene 1
Gene 2
Gene N

FIG. 4

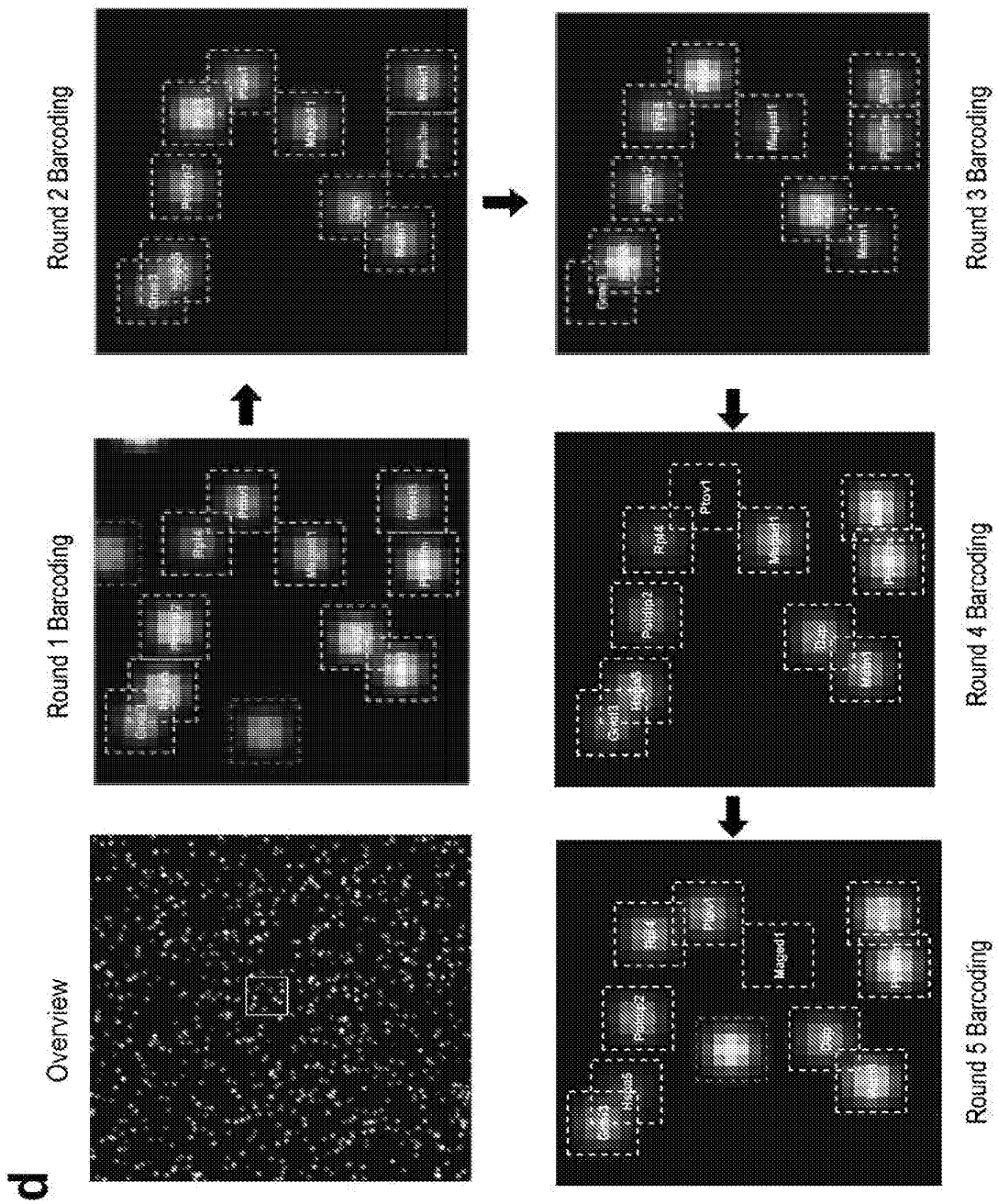


FIG. 4 (continued)

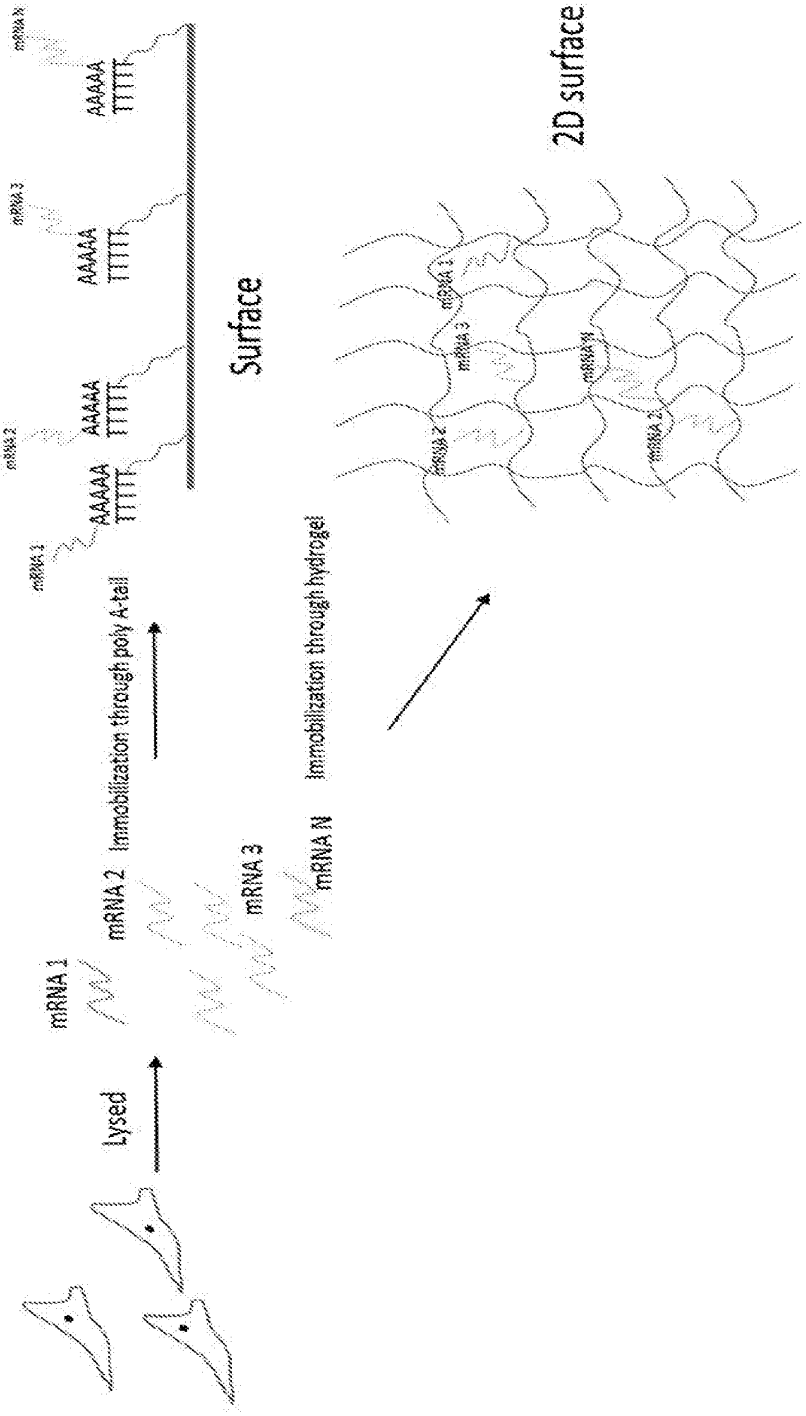

Lysed
mRNA 1
mRNA 2
mRNA 3
mRNA N
Immobilization through poly A-tail
Immobilization through hydrogel
Surface
2D surface
AAAAA
TTTTT

FIG. 5

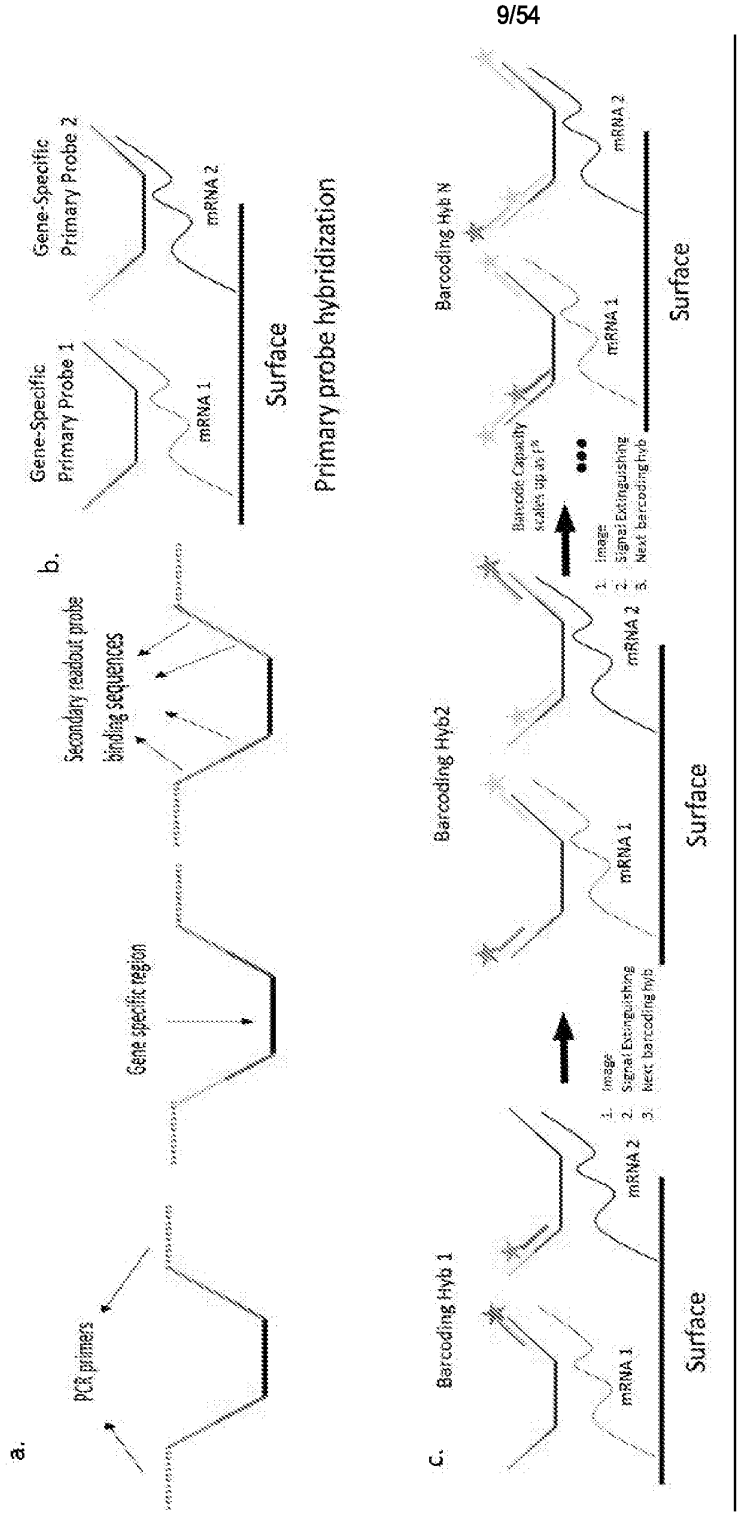

a.
PCR primers
Gene specific region
Secondary readout probe binding sequences
b.
Gene-Specific Primary Probe 1
Gene-Specific Primary Probe 2
mRNA 1
mRNA 2
Surface
Primary probe hybridization
c.
Barcoding Hyb 1
mRNA 1
mRNA 2
1. Image
2. Signal Extinguishing
3. Next barcoding hyb
Surface
Barcoding Hyb2
Barcoding Hyb N
Barcode Capacity scales up as P^n
Next barcoding hyb

FIG. 6

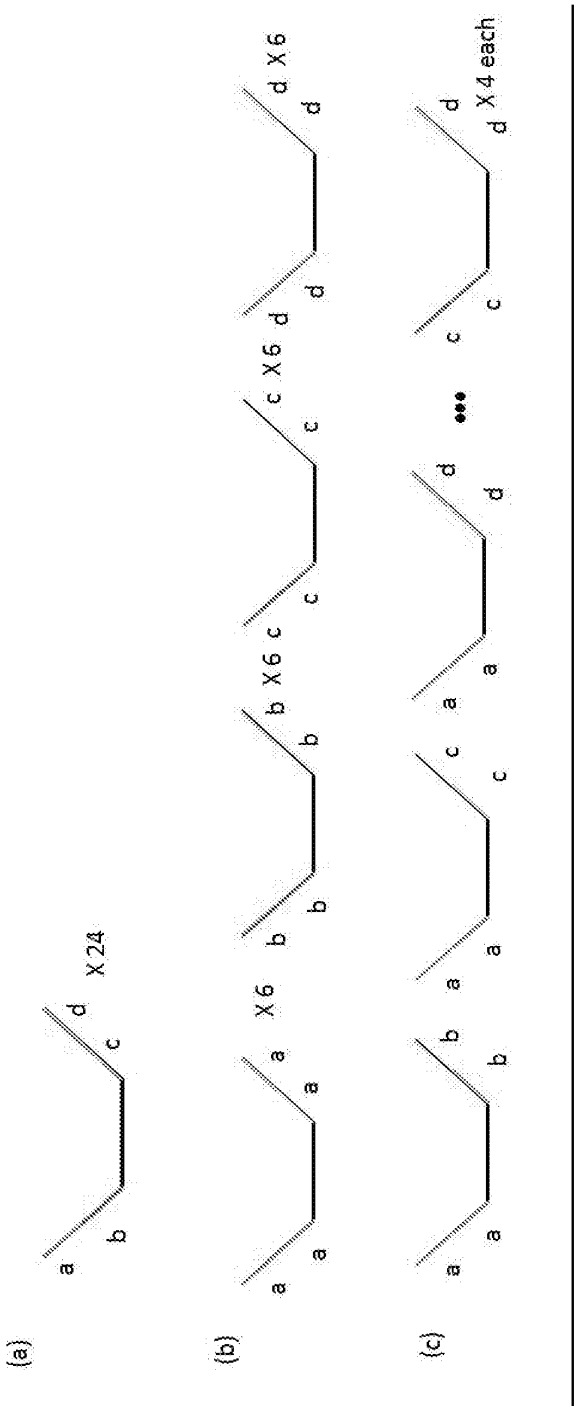

(a)
a
b
c
d
X 24
(b)
a
a
a
a
X 6
b
b
b
b
b X 6
c
c
c
c
c X 6
d
d
d
d
d X 6
(c)
a
a
a
b
b
a
a
c
c
a
a
d
d
c
c
d
d
X 4 each

FIG. 7

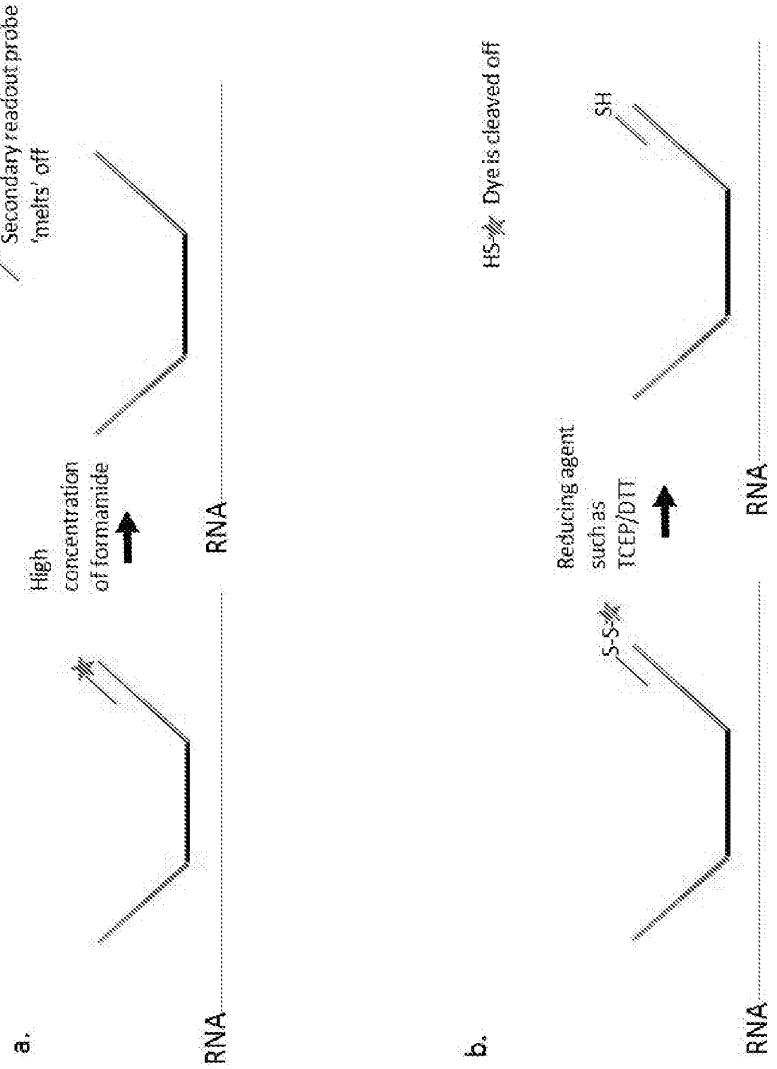

a.
RNA
High concentration of formamide
Secondary readout probe 'melts' off
RNA
b.
RNA
Reducing agent such as TCEP/DTT
S-S
RNA
HS
Dye is cleaved off
SH

FIG. 8

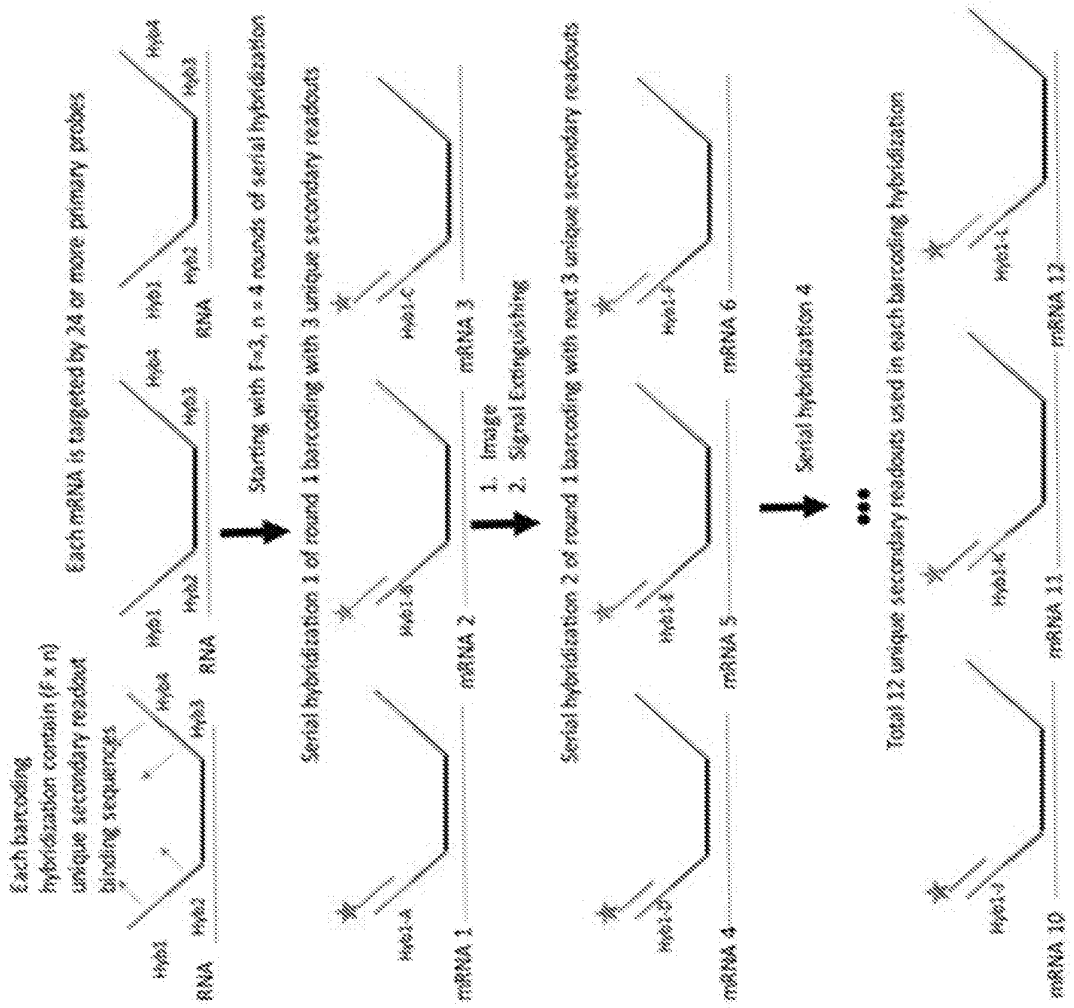

Each barcoding hybridization contain (r x n) unique secondary readout binding sequences
Each mRNA is targeted by 24 or more primary probes
Starting with r=3, n = 4 rounds of serial hybridization
Serial hybridization 1 of round 1 barcoding with 3 unique secondary readouts
1. Image
2. Signal Extinguishing
Serial hybridization 2 of round 1 barcoding with next 3 unique secondary readouts
Serial hybridization 4
Total 12 unique secondary readouts used in each barcoding hybridization
RNA
mRNA 1
mRNA 2
mRNA 3
mRNA 4
mRNA 5
mRNA 6
mRNA 10
mRNA 11
mRNA 12

FIG. 9

Serial Hyb	Barcoding Hyb1	Barcoding Hyb2	Barcoding Hyb3	Barcoding Hyb4
1	Readout 1 (Red)	Readout 13 (Red)	Readout 25 (Red)	Readout 37 (Red)
	Readout 2 (Green)	Readout 14 (Green)	Readout 26 (Green)	Readout 38 (Green)
	Readout 3 (Blue)	Readout 15 (Blue)	Readout 27 (Blue)	Readout 39 (Blue)
2	Readout 4 (Red)	Readout 16 (Red)	Readout 28 (Red)	Readout 40 (Red)
	Readout 5 (Green)	Readout 17 (Green)	Readout 29 (Green)	Readout 41 (Green)
	Readout 6 (Blue)	Readout 18 (Blue)	Readout 30 (Blue)	Readout 42 (Blue)
3	Readout 7 (Red)	Readout 19 (Red)	Readout 31 (Red)	Readout 43 (Red)
	Readout 8 (Green)	Readout 20 (Green)	Readout 32 (Green)	Readout 44 (Green)
	Readout 9 (Blue)	Readout 21 (Blue)	Readout 33 (Blue)	Readout 45 (Blue)
4	Readout 10 (Red)	Readout 22 (Red)	Readout 34 (Red)	Readout 46 (Red)
	Readout 11 (Green)	Readout 23 (Green)	Readout 35 (Green)	Readout 47 (Green)
	Readout 12 (Blue)	Readout 24 (Blue)	Readout 36 (Blue)	Readout 48 (Blue)

FIG. 10

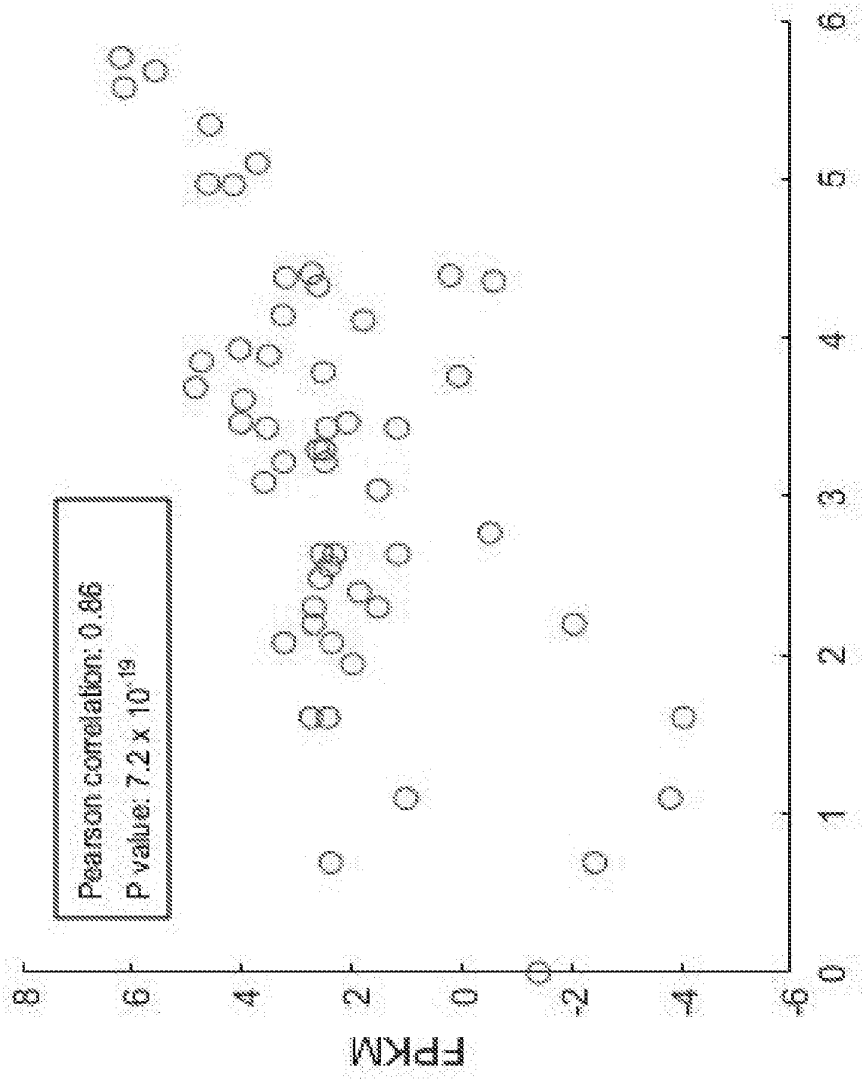

Pearson correlation: 0.86
P value: 7.2 x 10^-19
FPKM
Copy Number
8
6
4
2
0
-2
-4
-6
0
1
2
3
4
5
6

FIG. 11

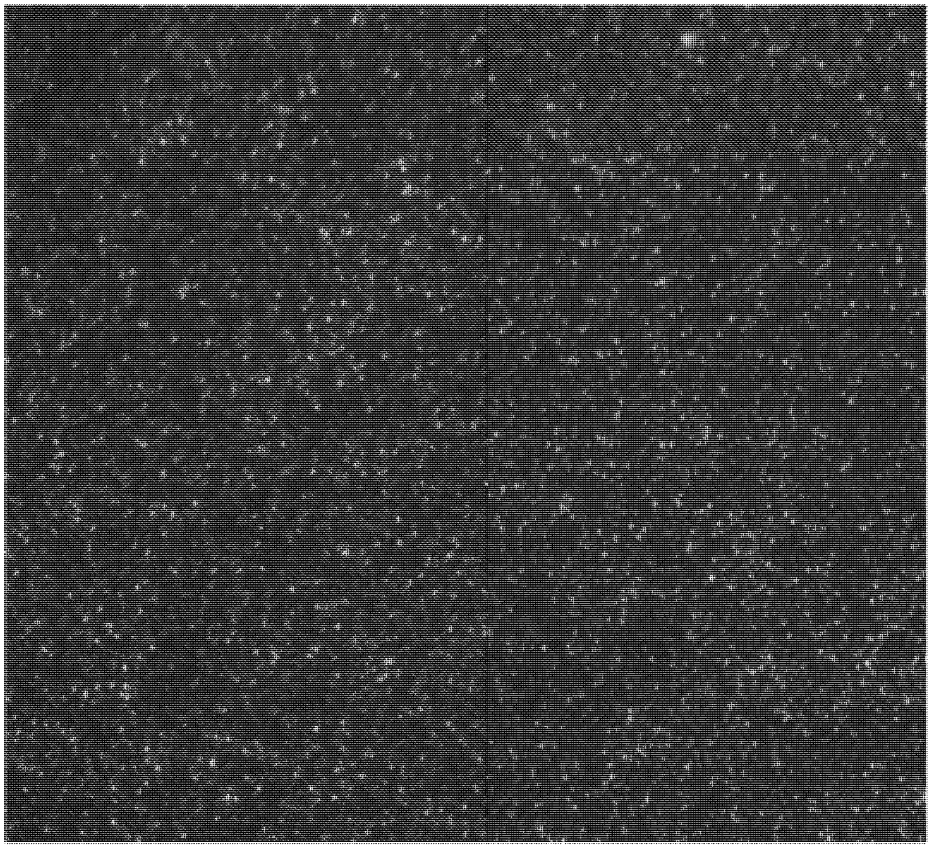


FIG. 12

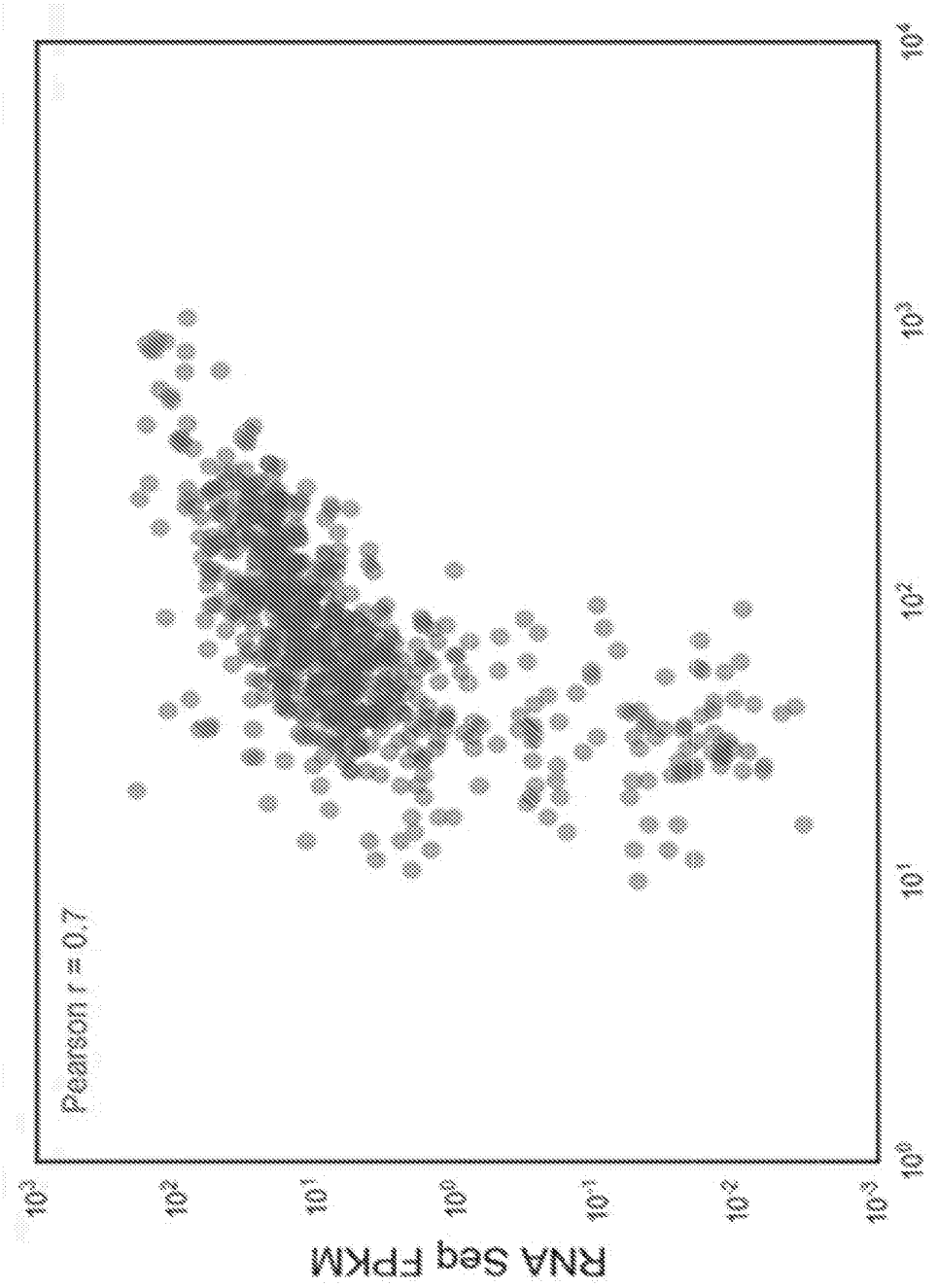

Pearson r = 0.7
RNA Seq FPKM
1000 genes SPOTs
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}

FIG. 13

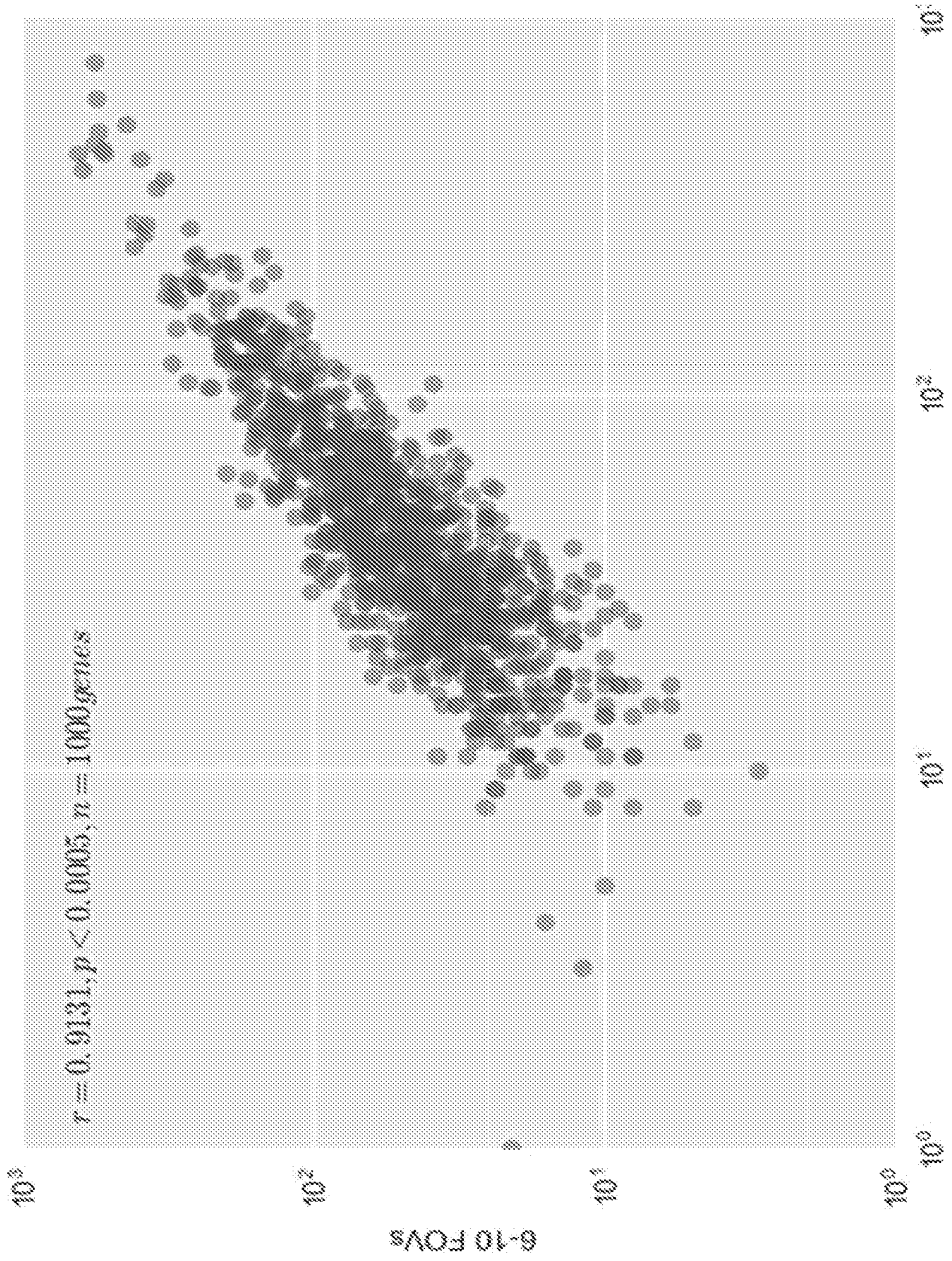


FIG. 14

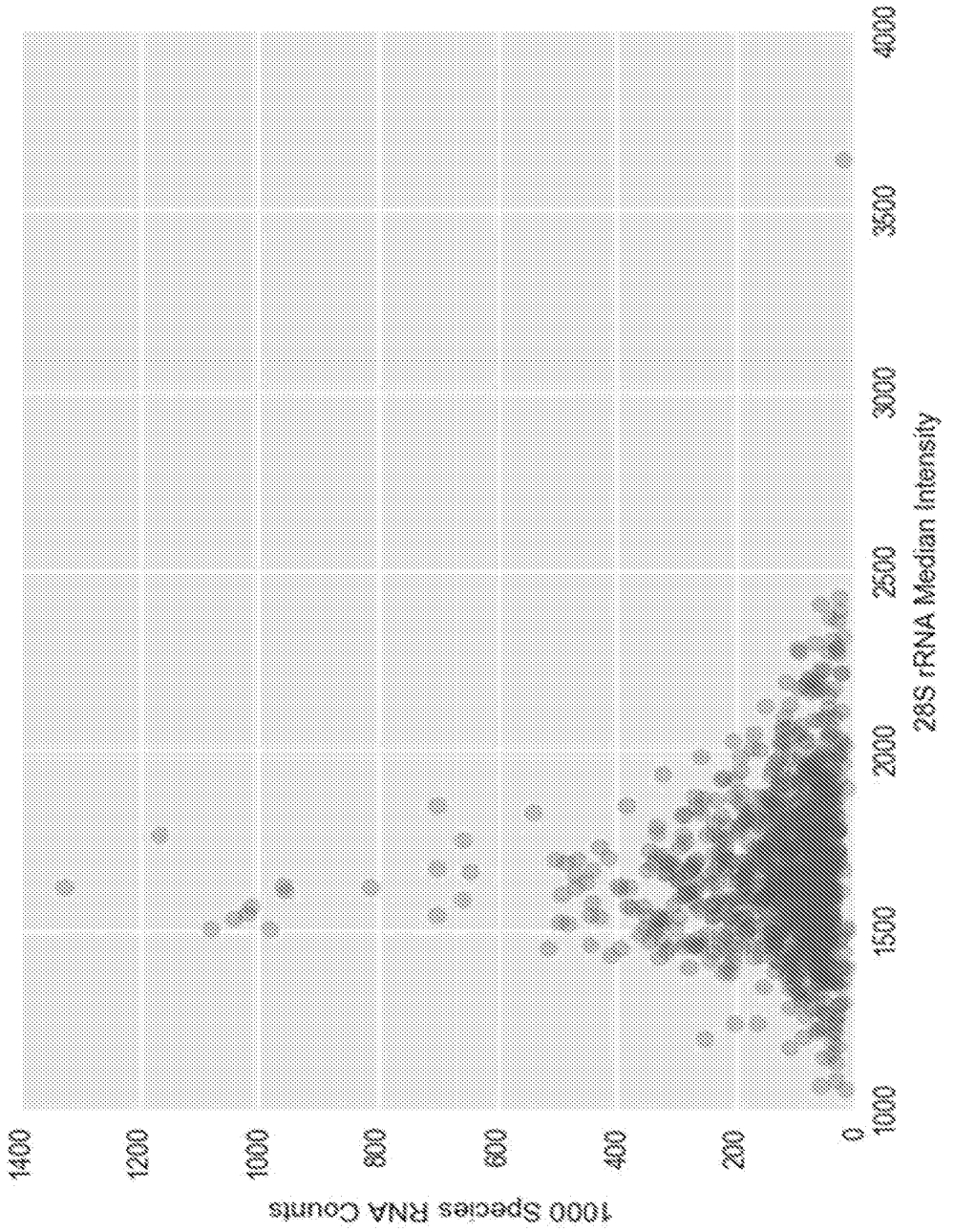

1000 Species RNA Counts
0
200
400
600
800
1000
1200
1400
28S rRNA Median Intensity
1000
1500
2000
2500
3000
3500
4000

FIG. 15

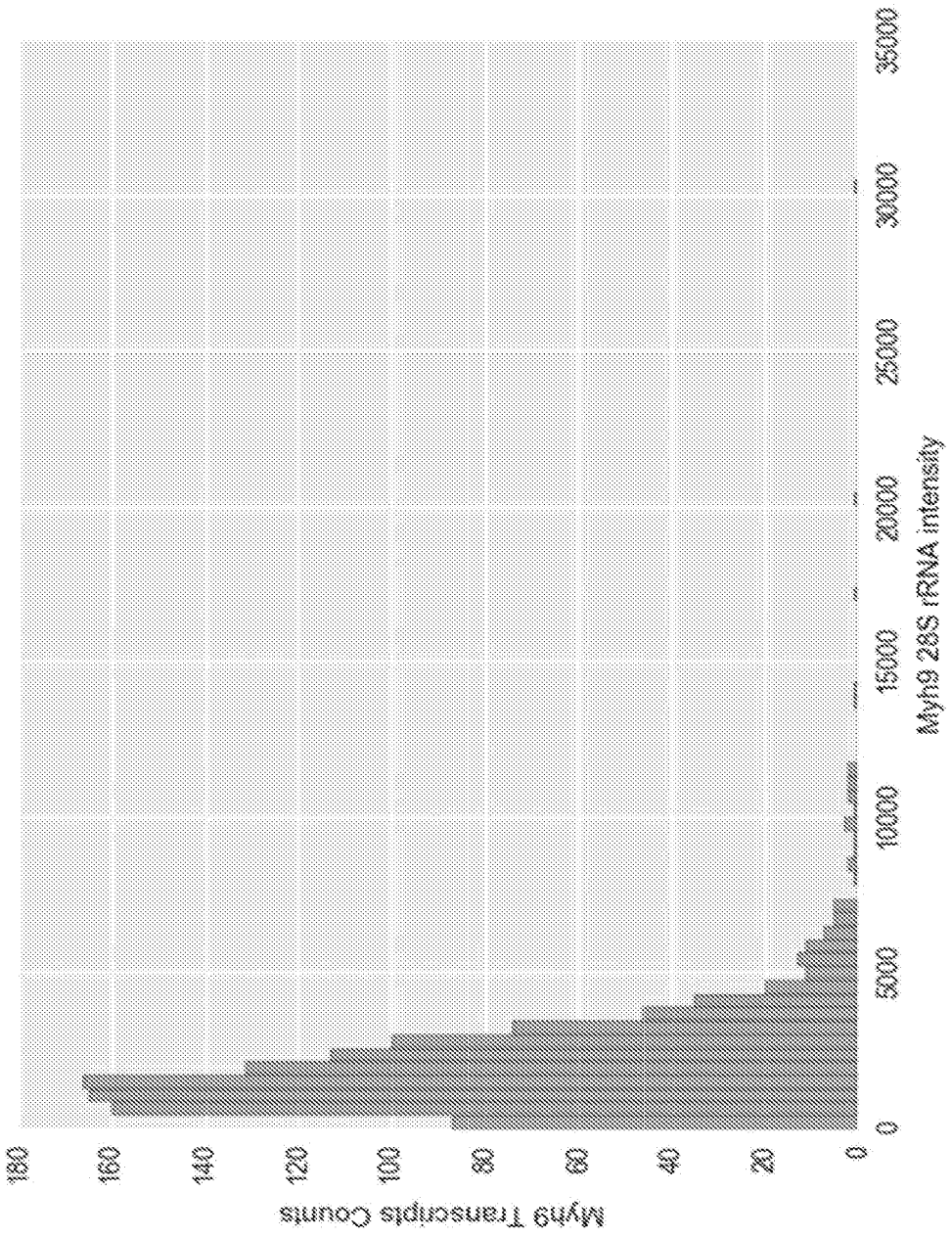

Myh9 Transcripts Counts
180
160
140
120
100
80
60
40
20
0
0
5000
10000
15000
20000
25000
30000
35000
Myh9 28S rRNA intensity

FIG. 16

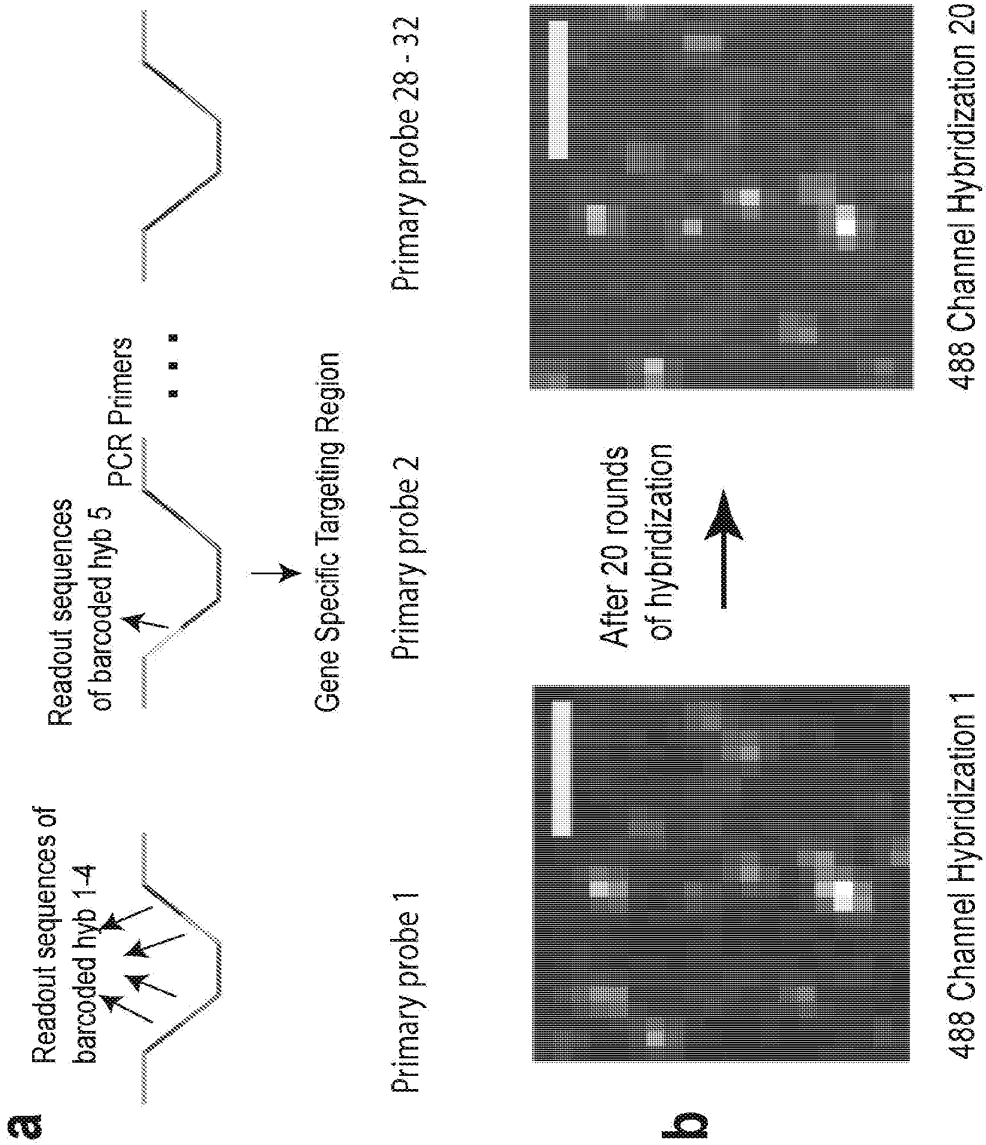
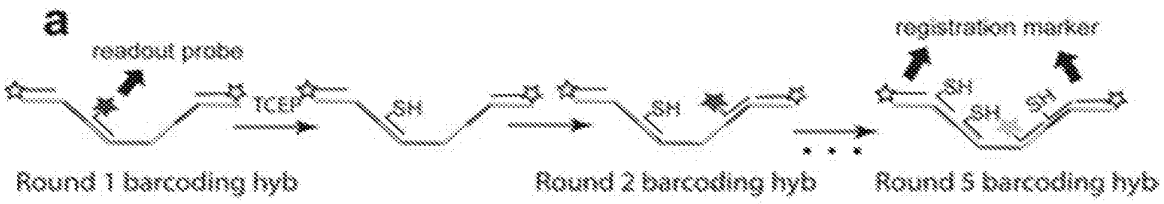

a
Readout sequences of barcoded hyb 1-4
Readout sequences of barcoded hyb 5
PCR Primers
Gene Specific Targeting Region
Primary probe 1
Primary probe 2
Primary probe 28 - 32
b
After 20 rounds of hybridization
488 Channel Hybridization 1
488 Channel Hybridization 20

FIG. 17


a
readout probe
TCEP
SH
SH
registration marker
SH
SH
SH
Round 1 barcoding hyb
Round 2 barcoding hyb
Round 5 barcoding hyb

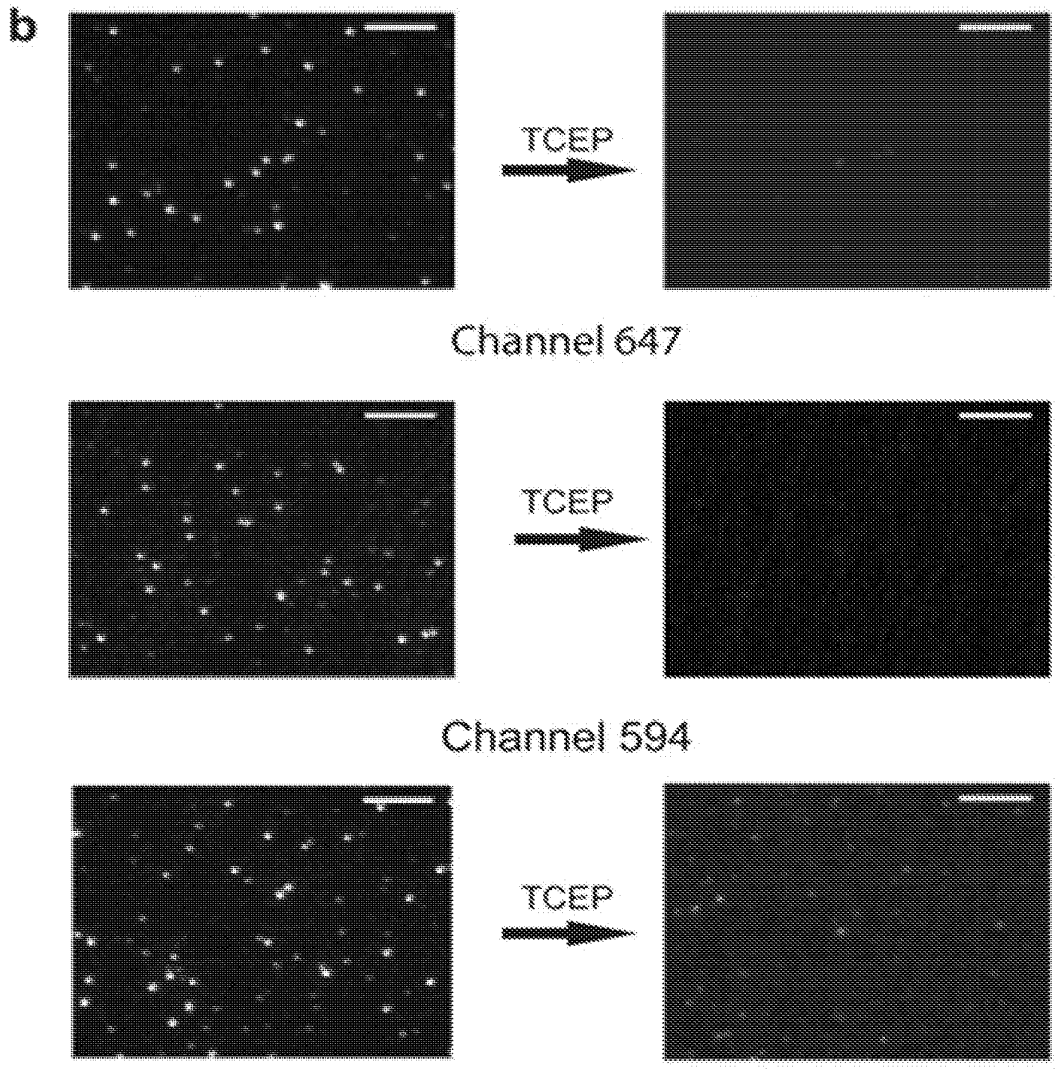

b
TCEP
Channel 647
TCEP
Channel 594
TCEP
Channel 532

FIG. 18

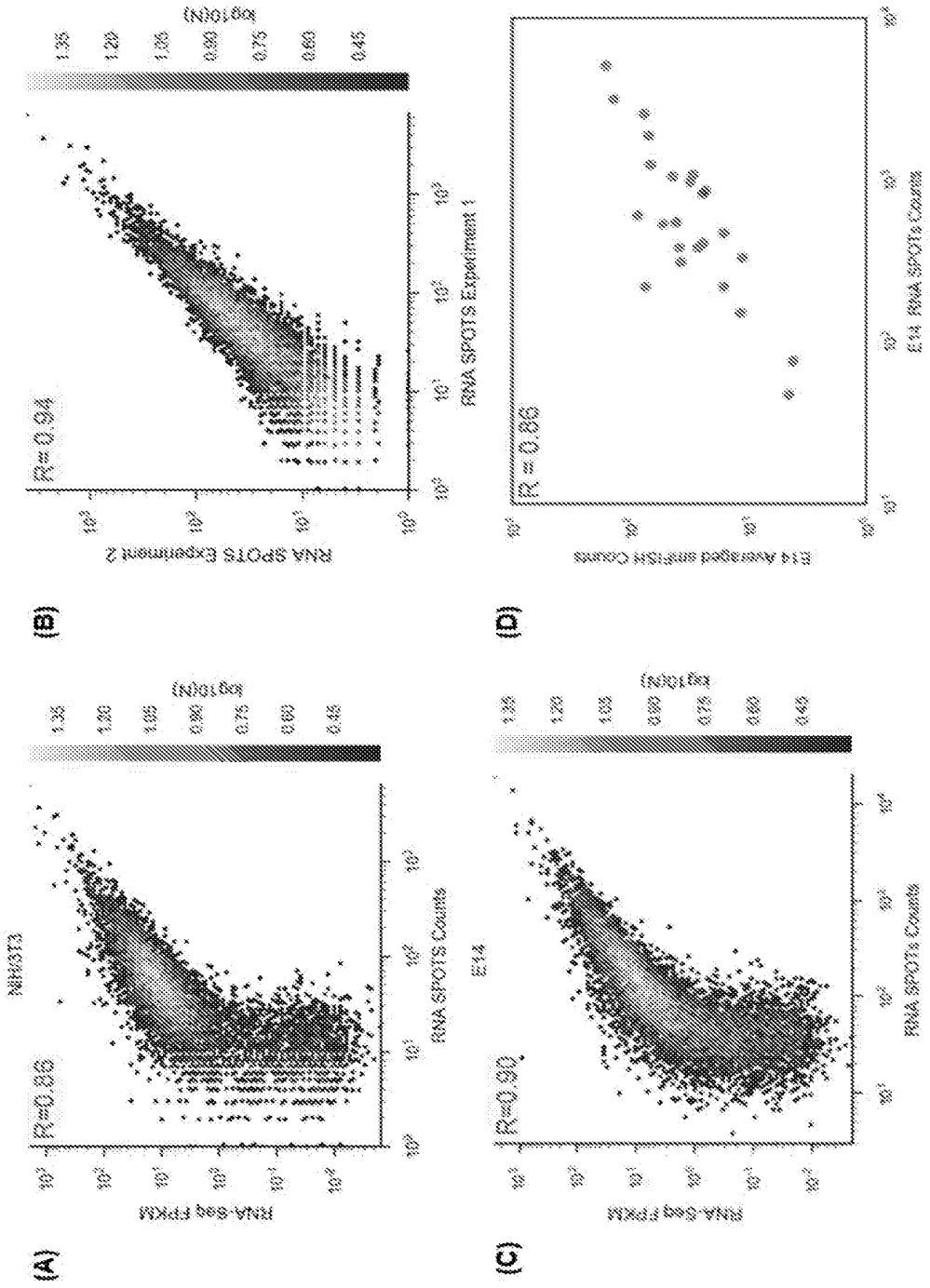

(A)
NIH/3T3
R=0.86
RNA-Seq FPKM
RNA SPOTS Counts
(B)
R= 0.94
RNA SPOTS Experiment 2
RNA SPOTS Experiment 1
(C)
E14
R=0.80
RNA-Seq FPKM
RNA SPOTs Counts
(D)
R = 0.86
E14 Averaged smFISH Counts
E14 RNA SPOTs Counts
log10(N)

FIG. 19

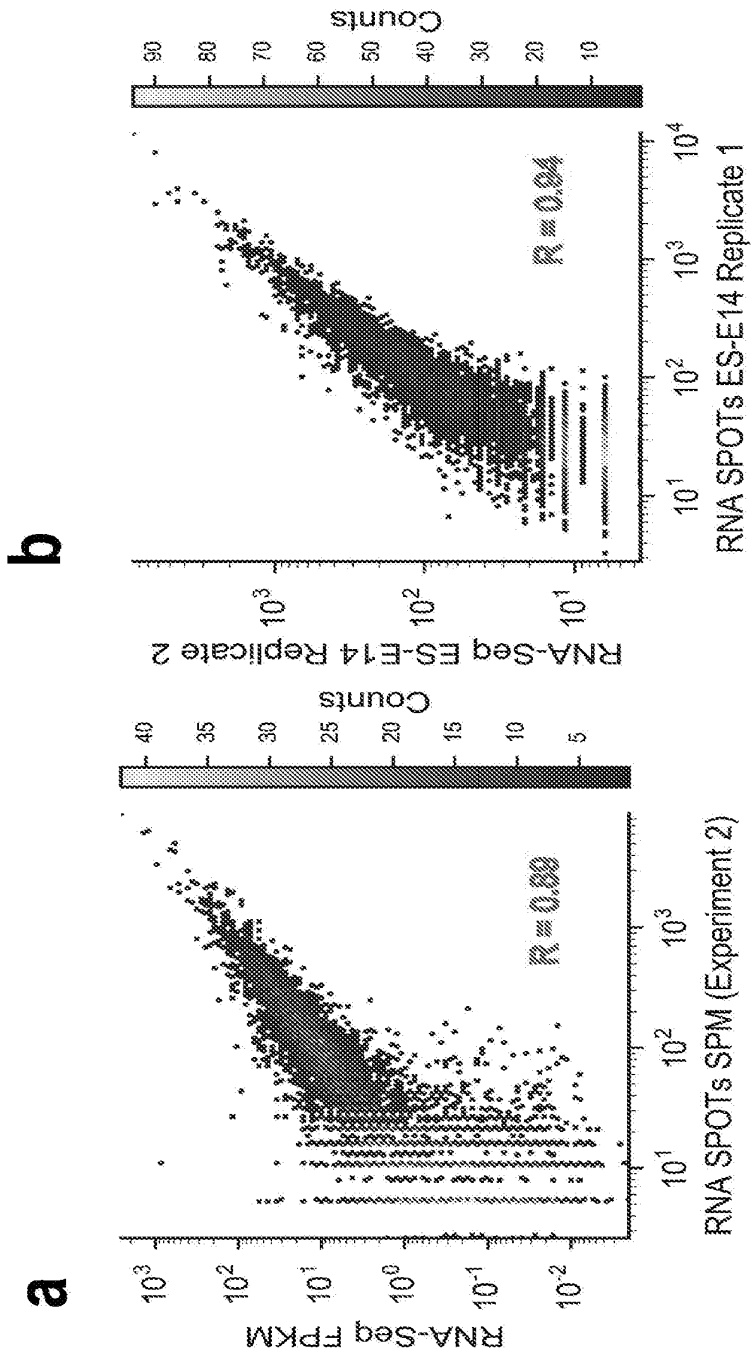

a
RNA-Seq FPKM
RNA SPOTs SPM (Experiment 2)
R = 0.89
Counts
b
RNA-Seq ES-E14 Replicate 2
RNA SPOTs ES-E14 Replicate 1
R = 0.94
Counts

FIG. 20

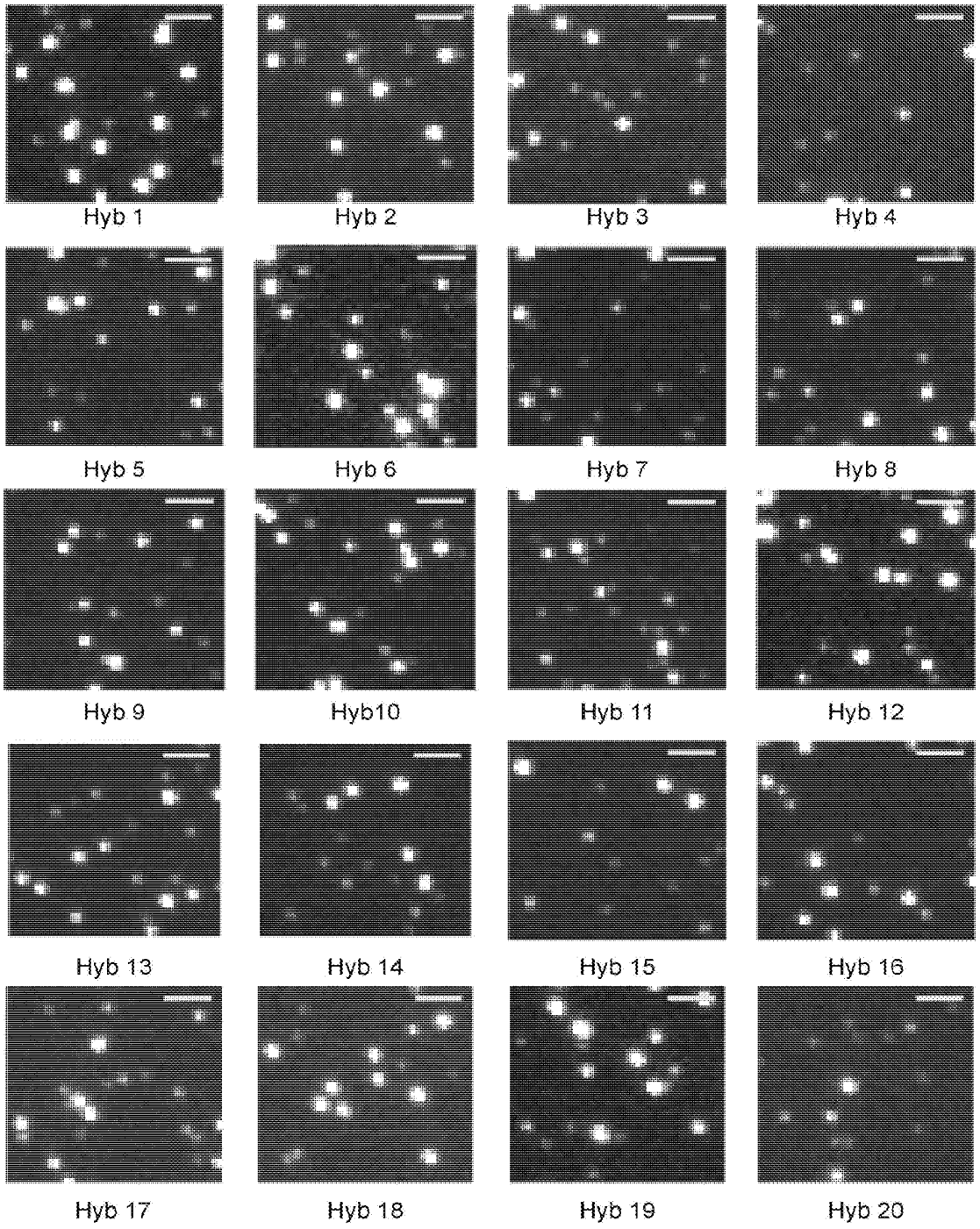

Hyb 1
Hyb 2
Hyb 3
Hyb 4
Hyb 5
Hyb 6
Hyb 7
Hyb 8
Hyb 9
Hyb10
Hyb 11
Hyb 12
Hyb 13
Hyb 14
Hyb 15
Hyb 16
Hyb 17
Hyb 18
Hyb 19
Hyb 20

FIG. 21

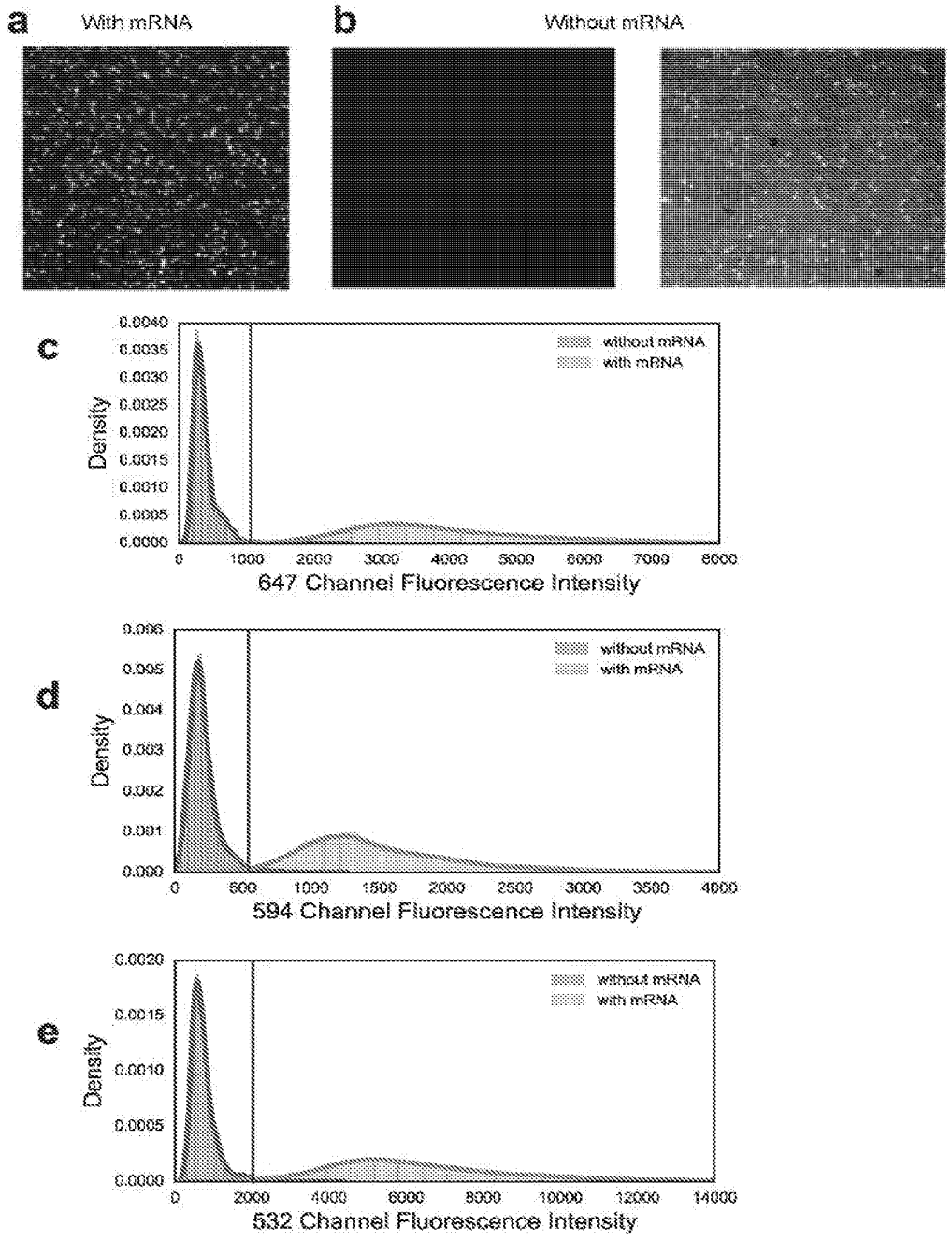

a
With mRNA
b
Without mRNA
c
Density
0.0040
0.0035
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
0
1000
2000
3000
4000
5000
6000
7000
8000
647 Channel Fluorescence Intensity
without mRNA
with mRNA
d
Density
0.006
0.005
0.004
0.003
0.002
0.001
0.000
0
500
1000
1500
2000
2500
3000
3500
4000
594 Channel Fluorescence Intensity
without mRNA
with mRNA
e
Density
0.0020
0.0015
0.0010
0.0005
0.0000
0
2000
4000
6000
8000
10000
12000
14000
532 Channel Fluorescence Intensity
without mRNA
with mRNA

FIG. 22

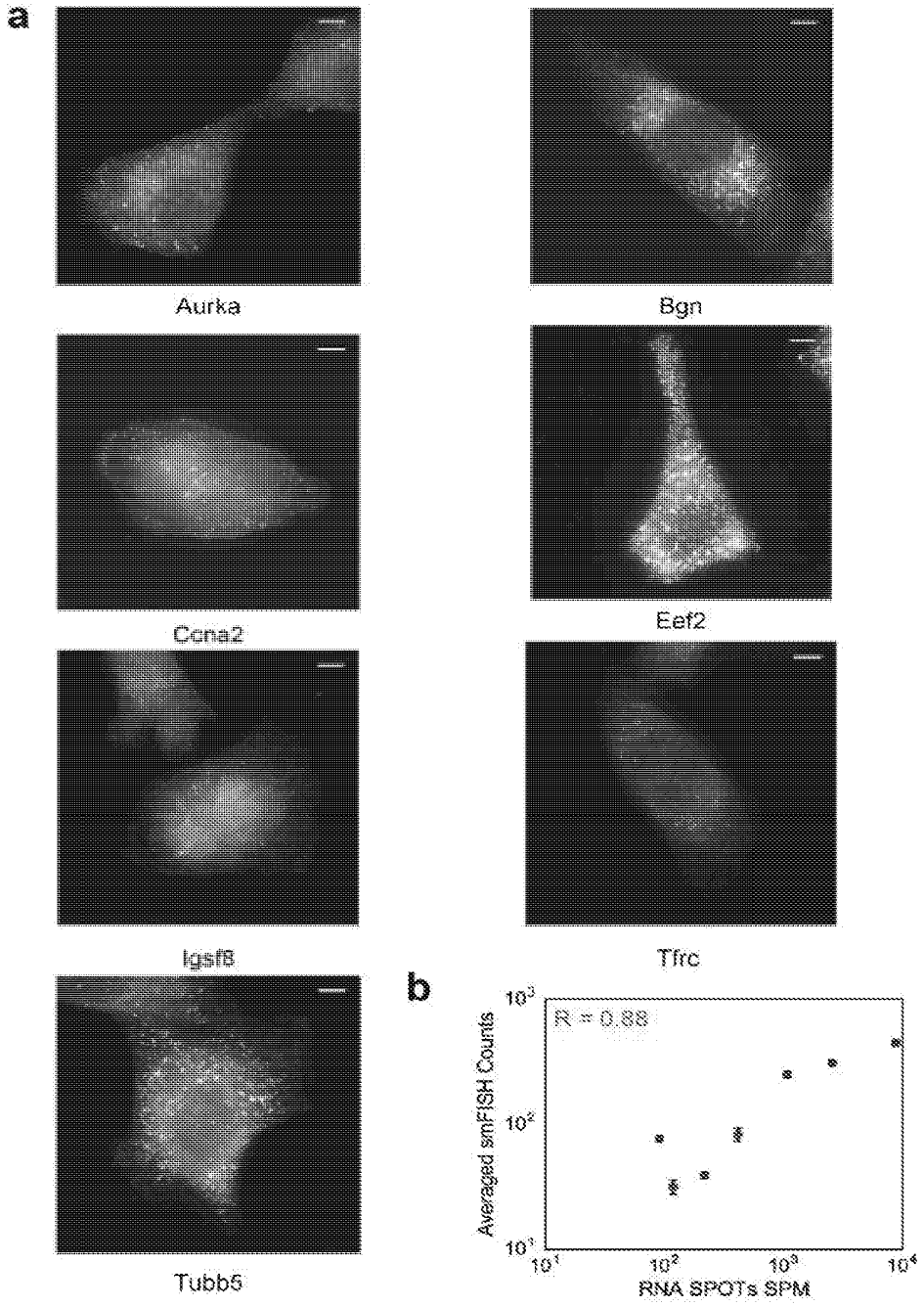

a
Aurka
Bgn
Ccna2
Eef2
Igsf8
Tfrc
Tubb5
b
R = 0.88
Averaged smFISH Counts
RNA SPOTs SPM

FIG. 23

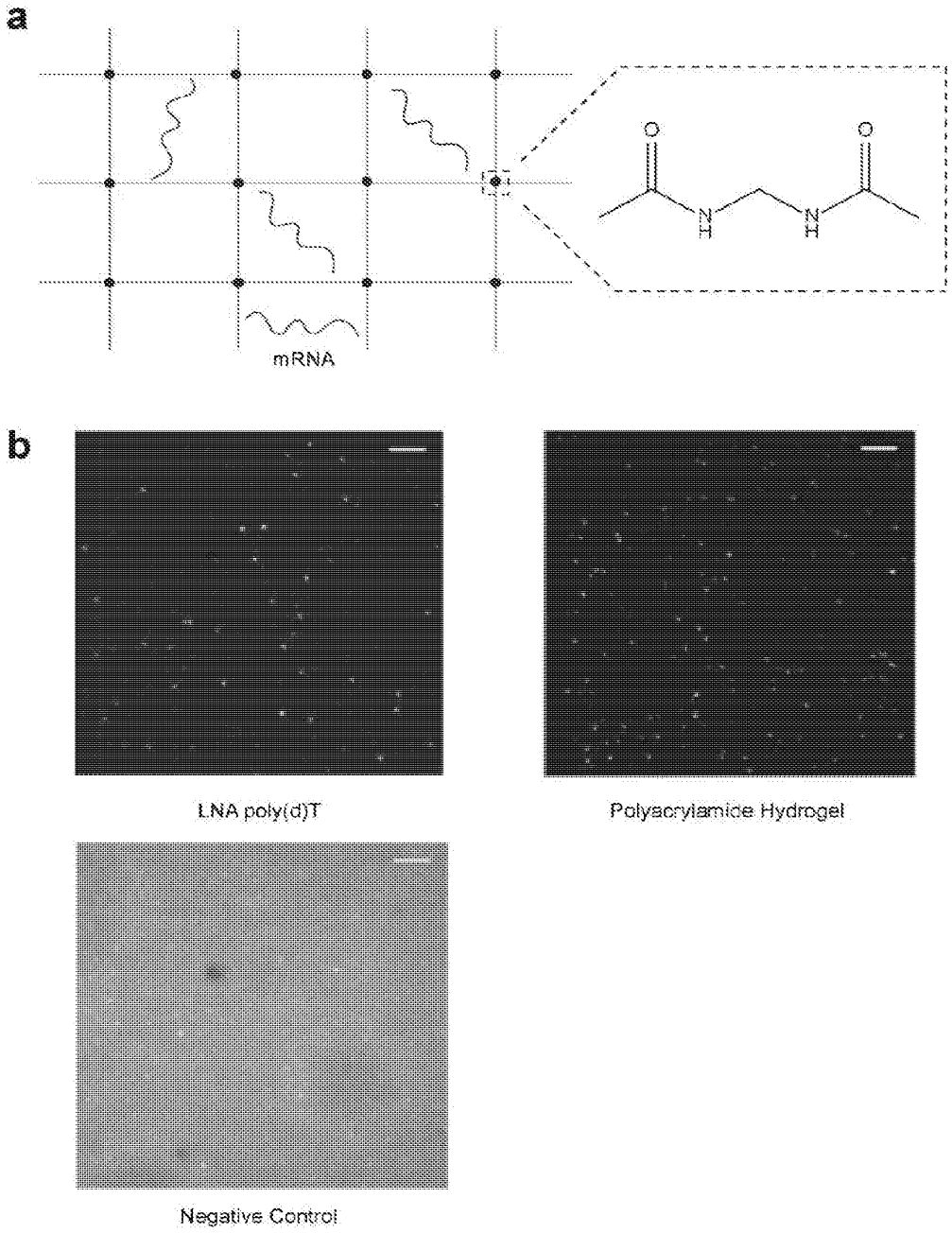

a
mRNA
b
LNA poly(d)T
Polyacrylamide Hydrogel
Negative Control

FIG. 24

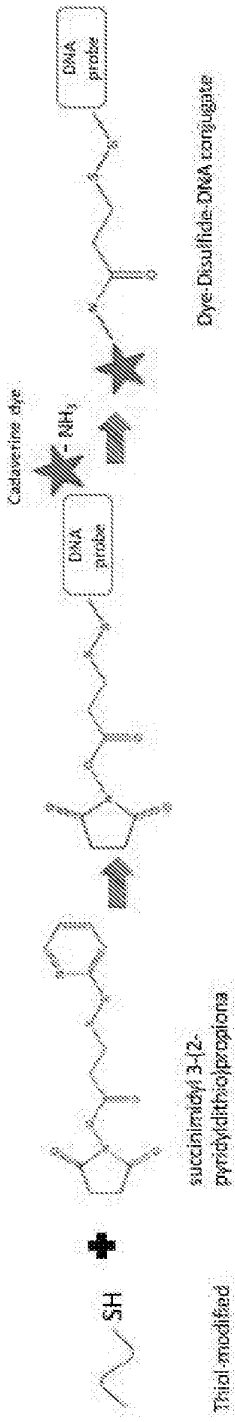

Thiol-modified DNA probes
SH
succinimidyl 3-(2-pyridyldithio)propionate (SPDP)
DNA probe
Cadaverine dye
NH2
DNA probe
Dye-Disulfide-DNA conjugate

FIG. 25

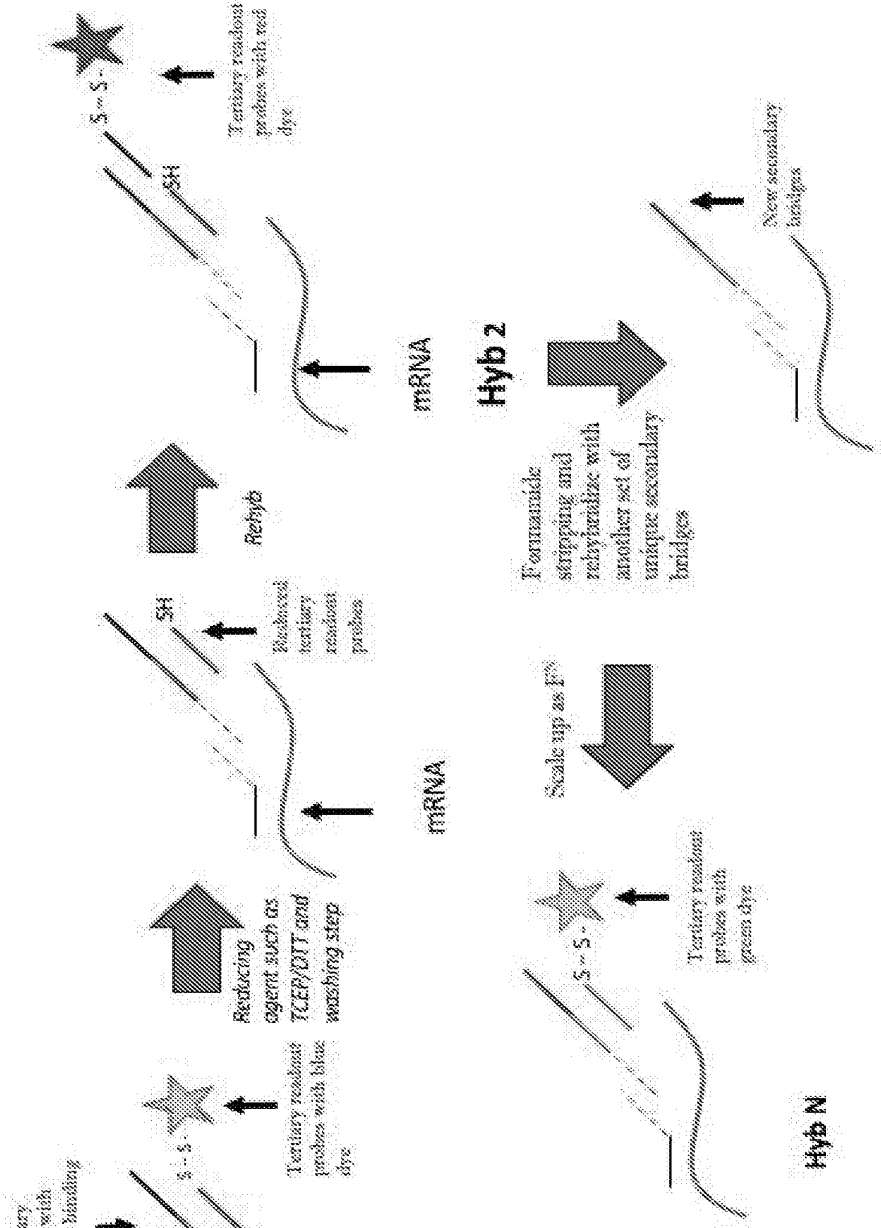

a
Gene specific primary probes
Secondary bridges with readout binding sites
S – S -
Tertiary readout probes with blue dye
Reducing agent such as TCEP/DTT and washing step
mRNA
Hyb 1
SH
Released tertiary readout probes
Rehyb
mRNA
Hyb 2
S – S -
Tertiary readout probes with red dye
Formamide stripping and rehybridize with another set of unique secondary bridges
New secondary bridges
Scale up as Fig
S – S -
Tertiary readout probes with green dye
Hyb N

FIG. 26

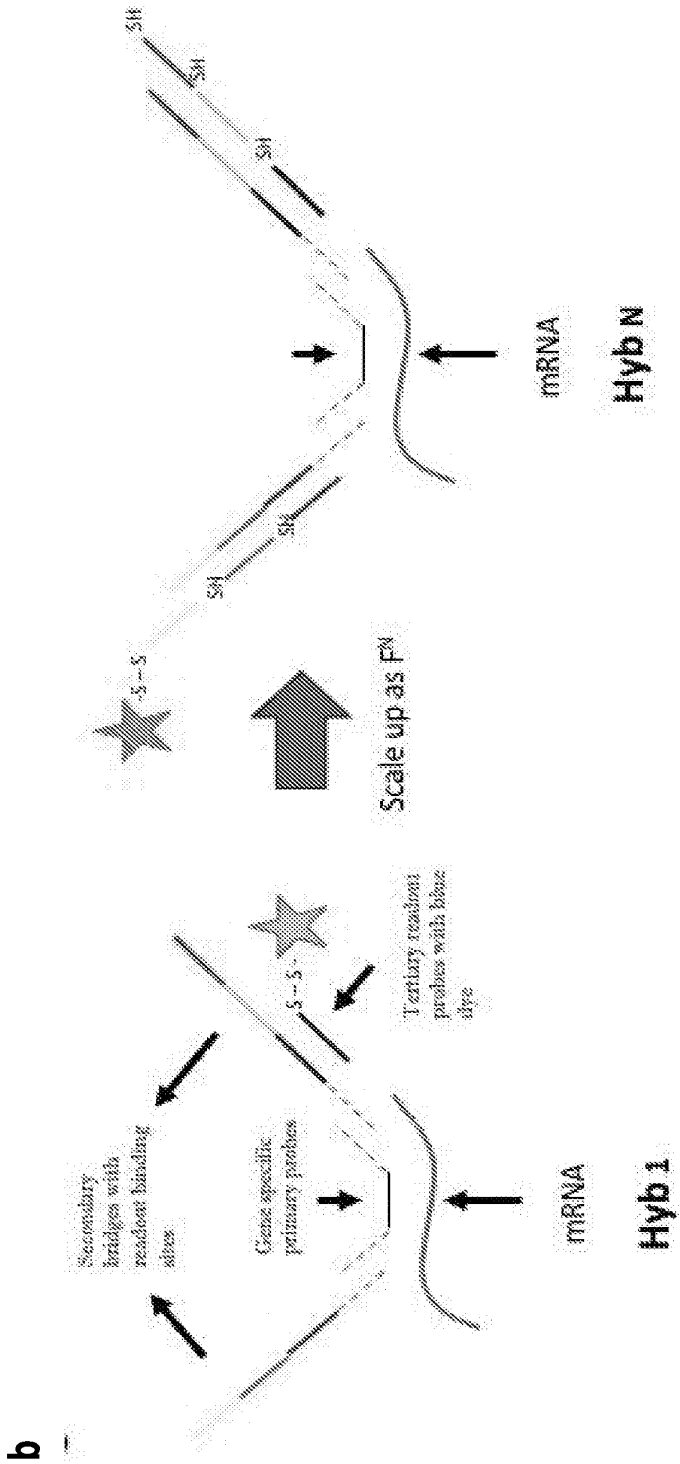


FIG. 26 (continued)

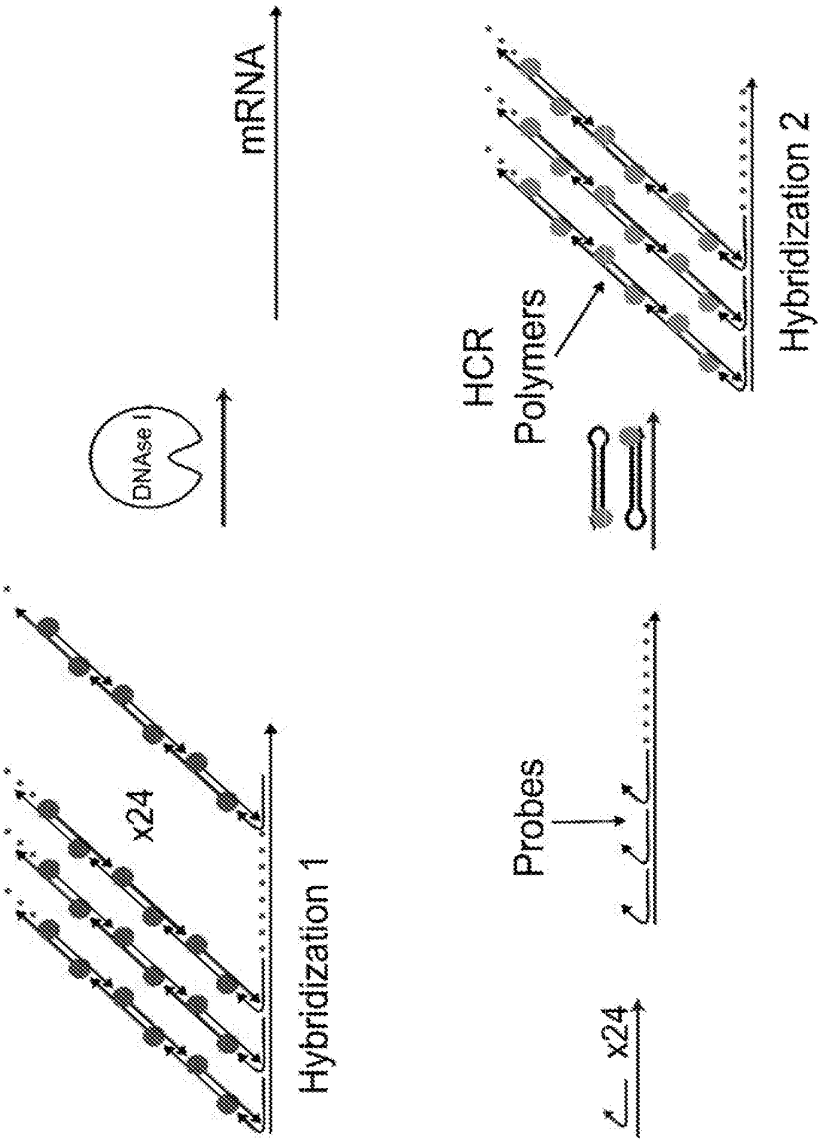


FIG. 27A (Prior Art)

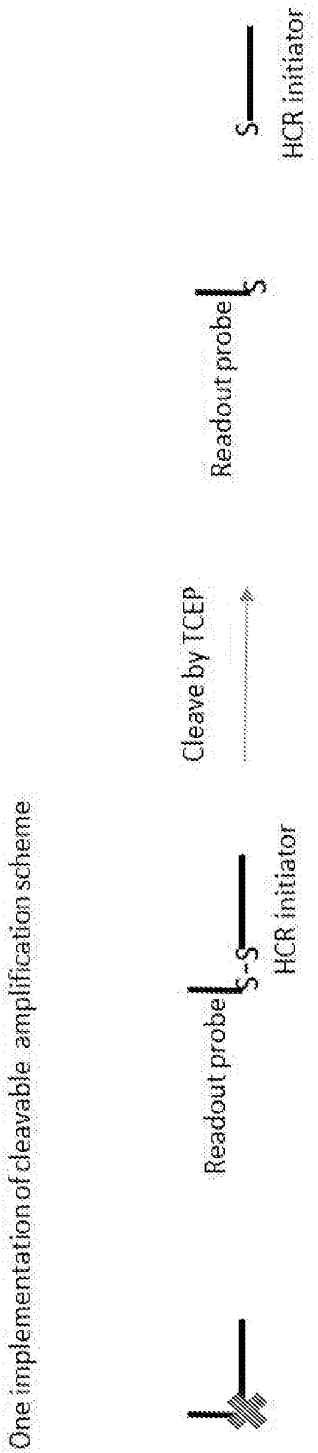

One implementation of cleavable amplification scheme
Readout probe
S-S
HCR initiator
Cleave by TCEP
Readout probe
S
S
HCR initiator

FIG. 27B

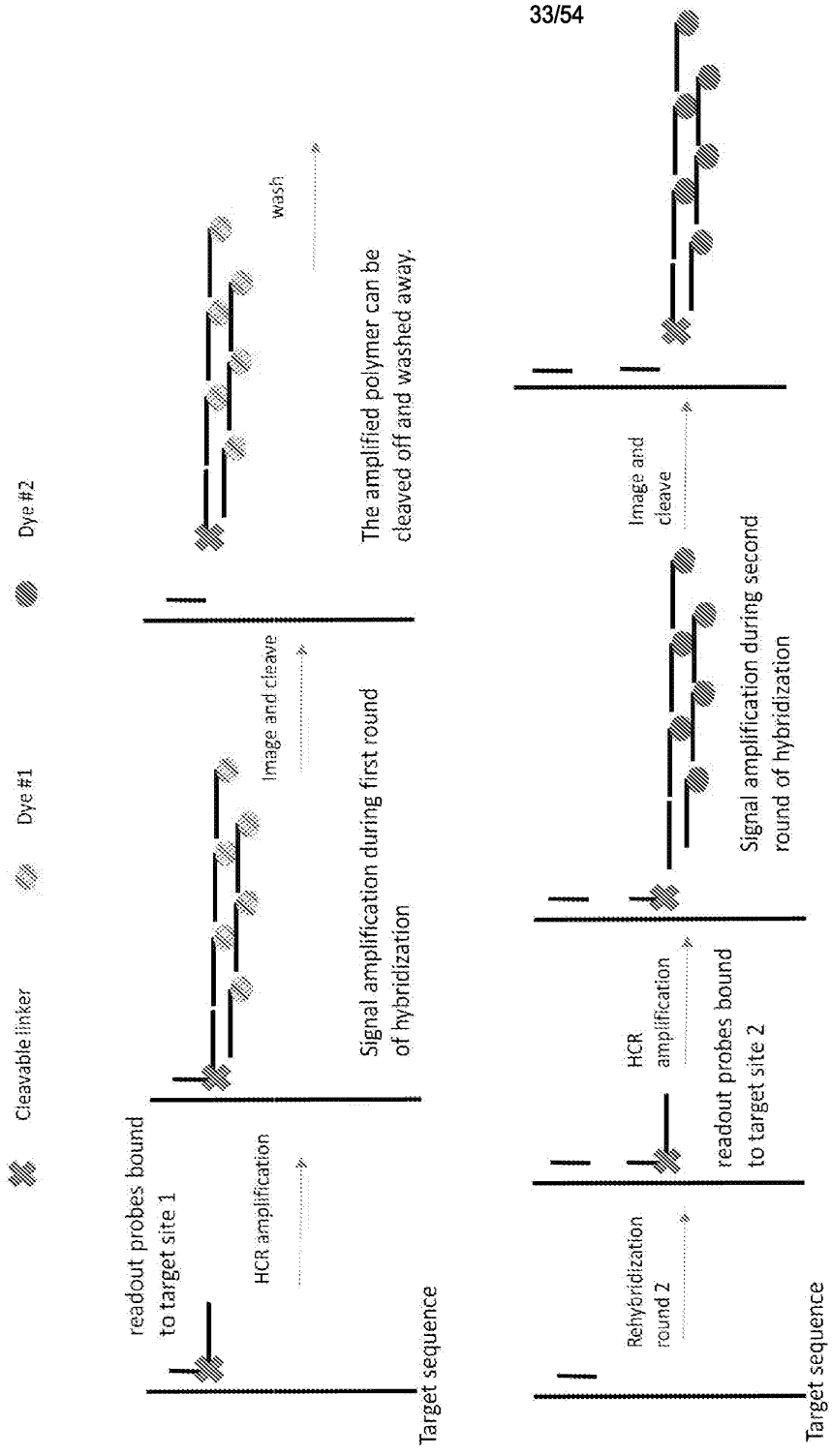

Cleavable linker
Dye #1
Dye #2
readout probes bound to target site 1
HCR amplification
Target sequence
Signal amplification during first round of hybridization
Image and cleave
wash
The amplified polymer can be cleaved off and washed away.
Rehybridization round 2
HCR amplification
readout probes bound to target site 2
Target sequence
Signal amplification during second round of hybridization
Image and cleave

FIG. 27C

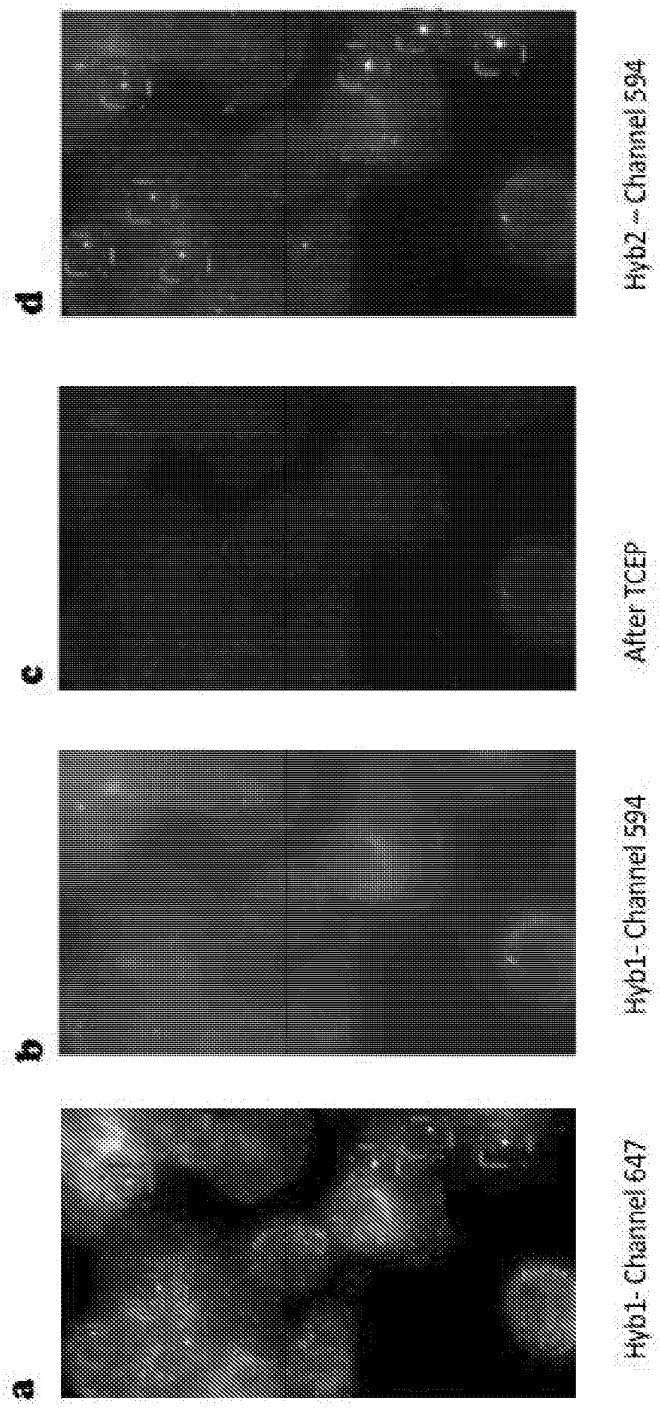

a
Hyb1- Channel 647
b
Hyb1- Channel 594
c
After TCEP cleavage - Channel 647
d
Hyb2 – Channel 594

FIG. 28

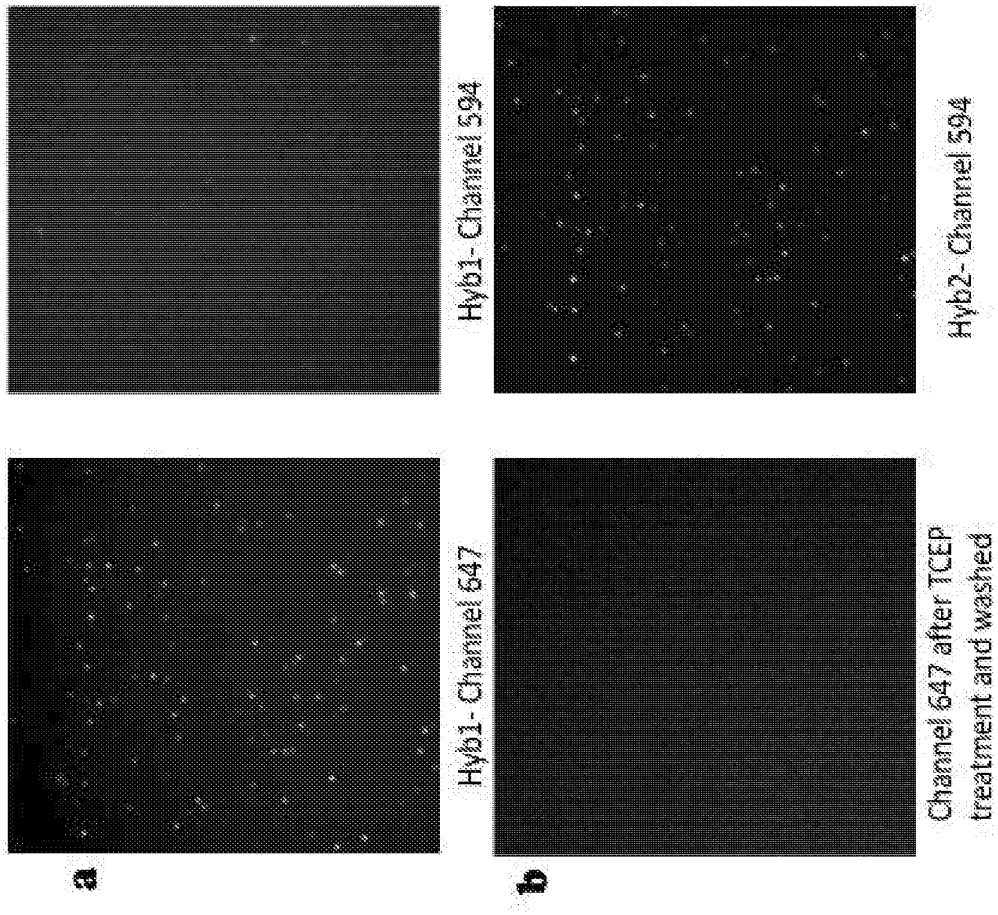

a
Hyb1- Channel 647
Hyb1- Channel 594
b
Channel 647 after TCEP treatment and washed
Hyb2- Channel 594

FIG. 29

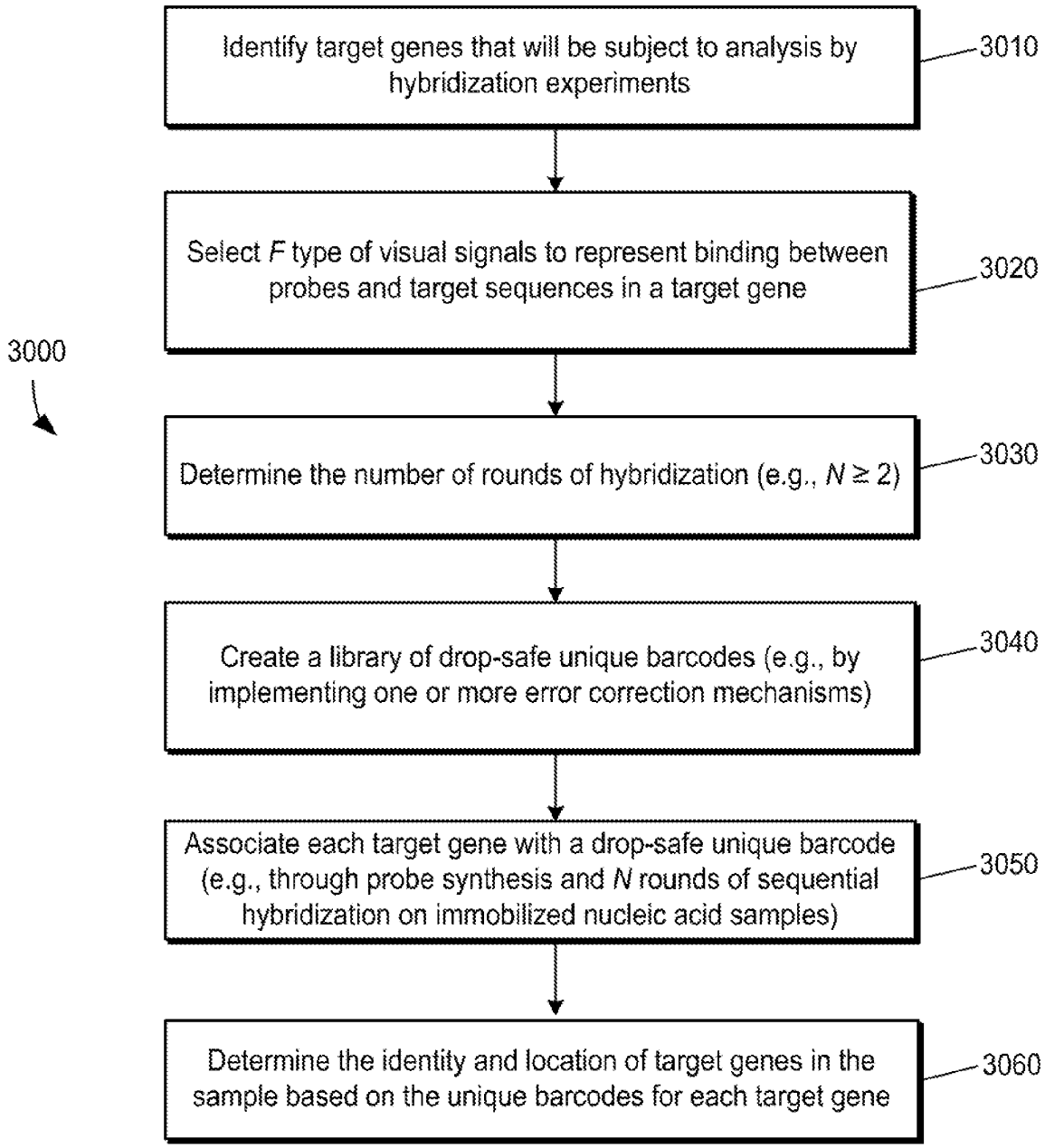

3000
Identify target genes that will be subject to analysis by hybridization experiments
3010
Select F type of visual signals to represent binding between probes and target sequences in a target gene
3020
Determine the number of rounds of hybridization (e.g., N ≥ 2)
3030
Create a library of drop-safe unique barcodes (e.g., by implementing one or more error correction mechanisms)
3040
Associate each target gene with a drop-safe unique barcode (e.g., through probe synthesis and N rounds of sequential hybridization on immobilized nucleic acid samples)
3050
Determine the identity and location of target genes in the sample based on the unique barcodes for each target gene
3060

FIG. 30

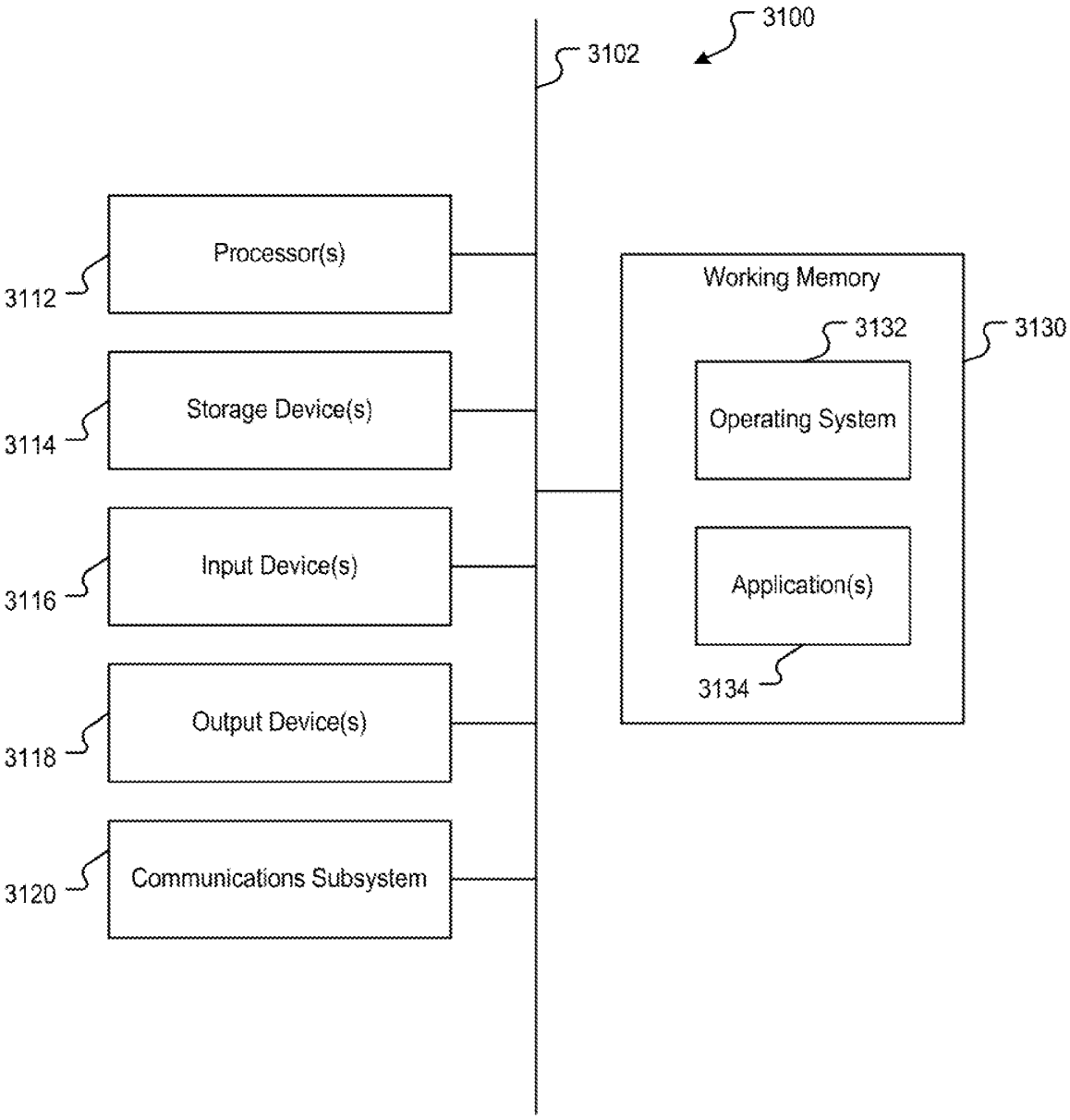

3100
3102
Processor(s)
3112
Storage Device(s)
3114
Input Device(s)
3116
Output Device(s)
3118
Communications Subsystem
3120
Working Memory
3130
3132
Operating System
Application(s)
3134

FIG. 31

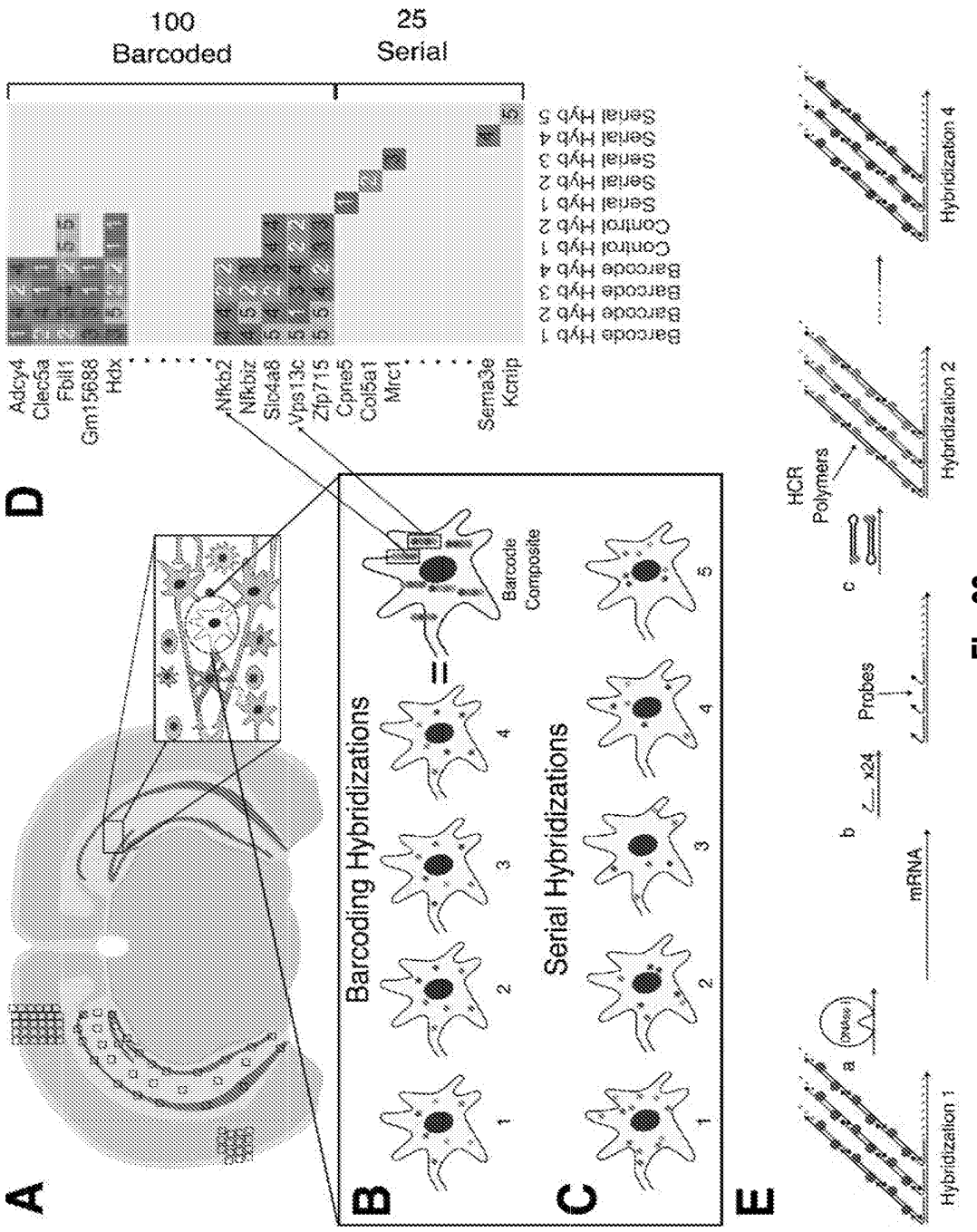

A
B
Barcoding Hybridizations
Barcode Composite
C
Serial Hybridizations
D
100 Barcoded
25 Serial
Adcy4
Clec5a
Fbil1
Gm15688
Hdx
Nfkb2
Nfkbiz
Slc4a8
Vps13c
Zfp715
Cpne5
Col5a1
Mrc1
Sema3e
Kcnip
Barcode Hyb 1
Barcode Hyb 2
Barcode Hyb 3
Barcode Hyb 4
Control Hyb 1
Control Hyb 2
Serial Hyb 1
Serial Hyb 2
Serial Hyb 3
Serial Hyb 4
Serial Hyb 5
E
Hybridization 1
mRNA
x24
Probes
HCR Polymers
Hybridization 2
Hybridization 4

Fig. 32

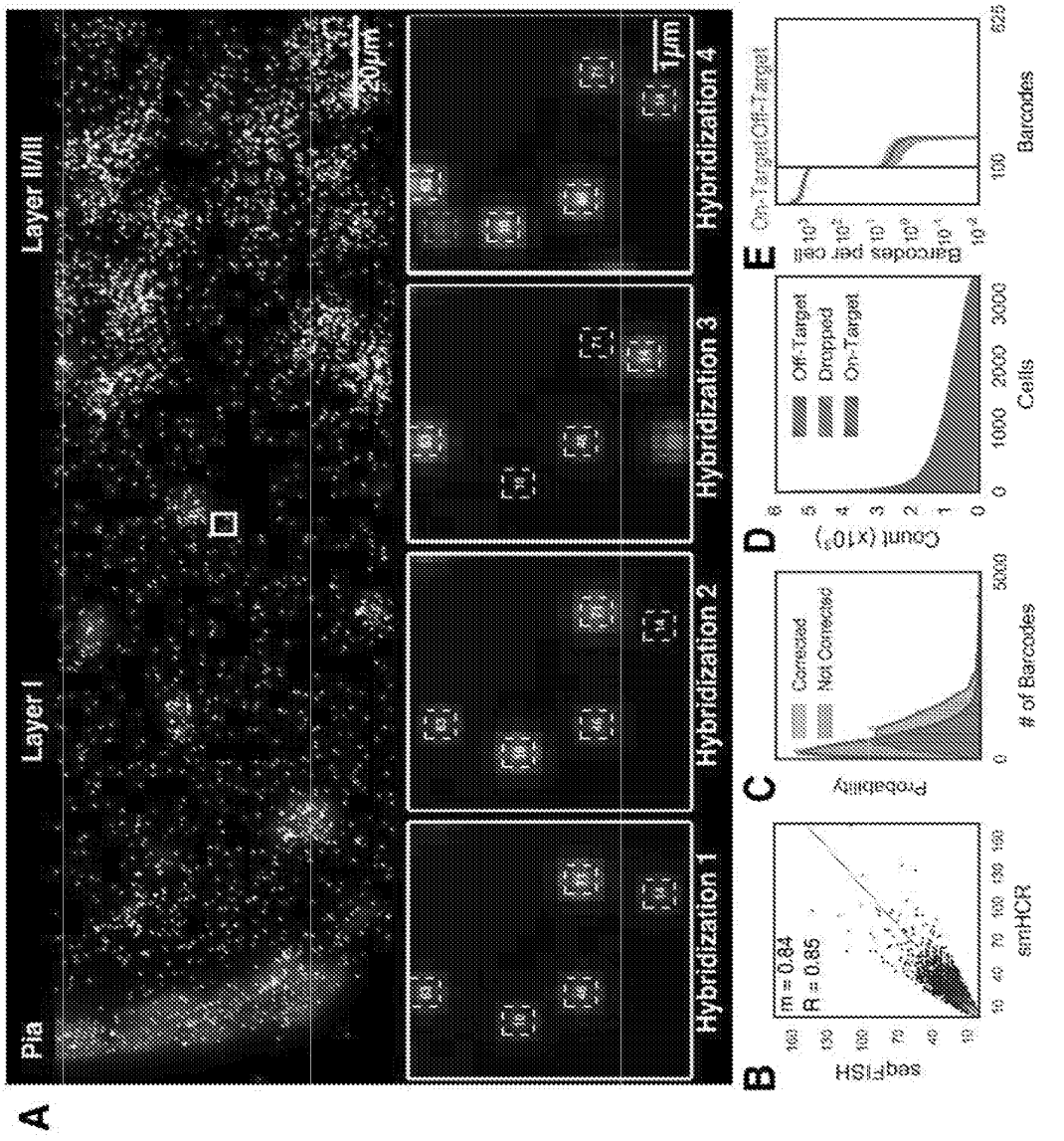

A
Pia
Layer I
Layer II/III
20μm
1μm
Hybridization 1
Hybridization 2
Hybridization 3
Hybridization 4
B
seqFISH
smHCR
m = 0.84
R = 0.85
C
Probability
Corrected
Not Corrected
of Barcodes
D
Count (x10^6)
Off-Target
Dropped
On-Target
Cells
E
On-Target/Off-Target
Barcodes per cell
Barcodes

Fig. 33

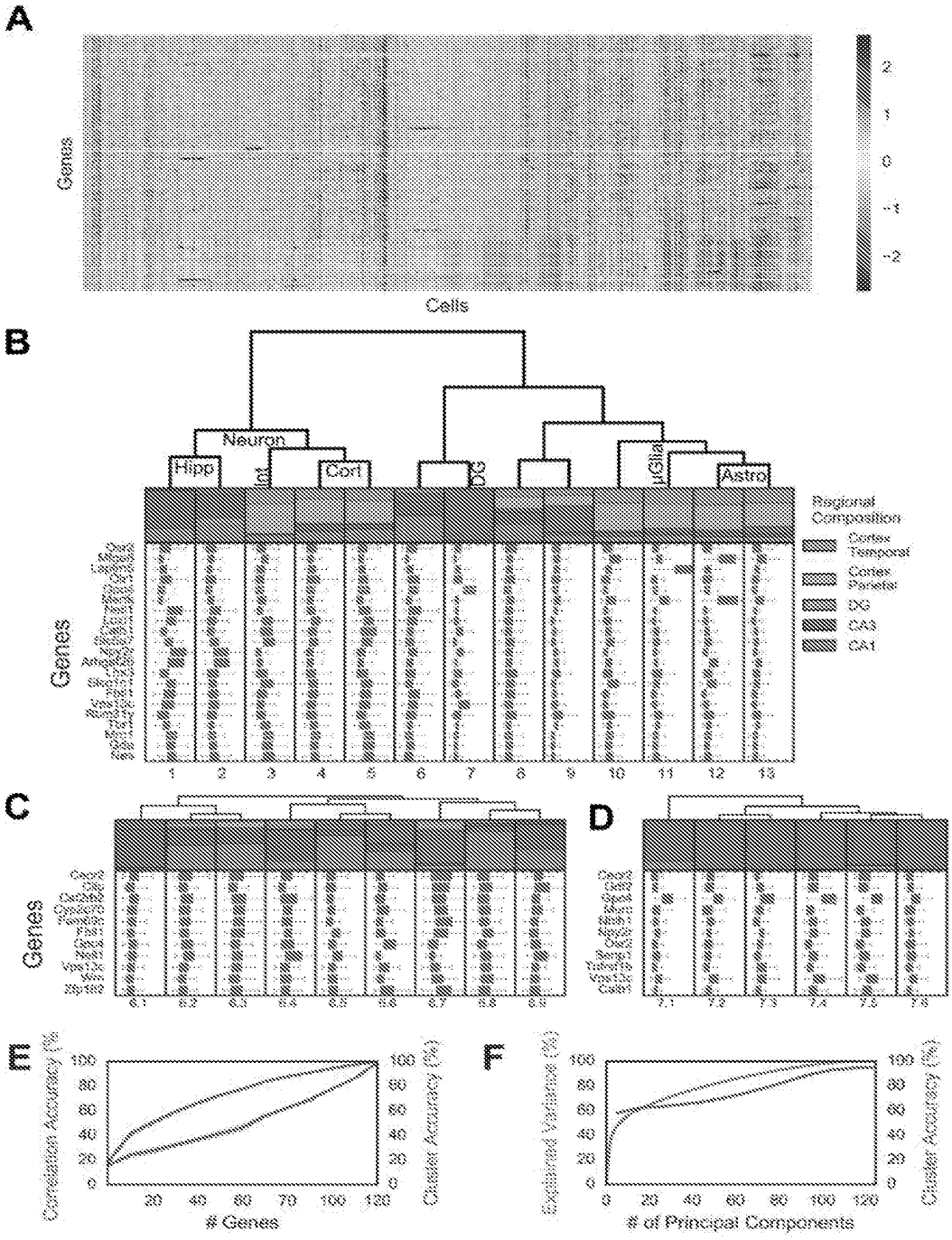

A
Genes
Cells
2
1
0
-1
-2
B
Neuron
Hipp
Int
Cort
OL
nGlia
Astro
Regional Composition
Cortex Temporal
Cortex Parietal
DG
CA3
CA1
Genes
1 2 3 4 5 6 7 8 9 10 11 12 13
C
D
Genes
E
Correlation Accuracy (%)
Cluster Accuracy (%)
0 20 40 60 80 100
20 40 60 70 100 120
Genes
F
Explained Variance (%)
Cluster Accuracy (%)
0 20 40 60 80 100 120
of Principal Components

Fig. 34

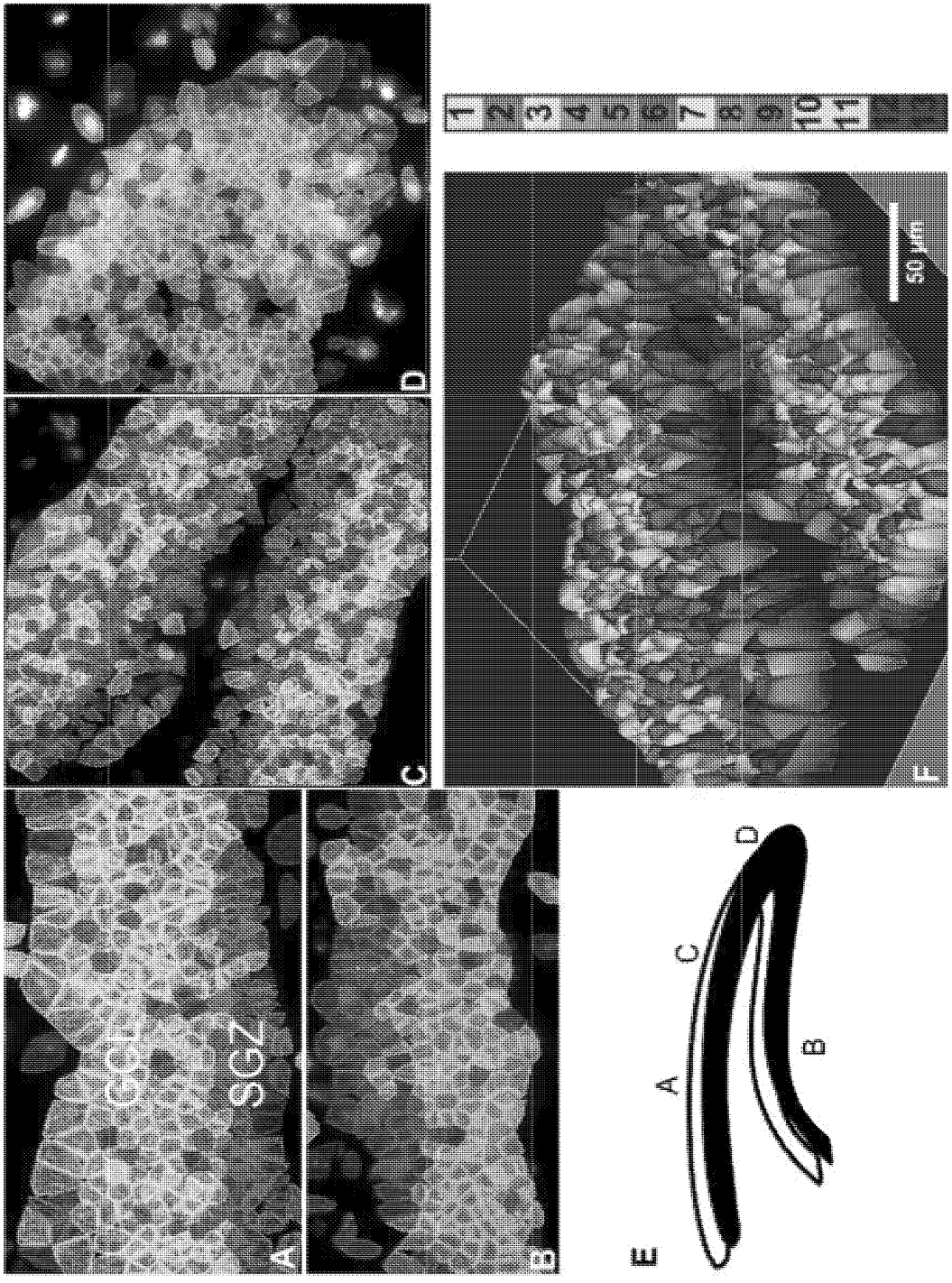


Fig. 35

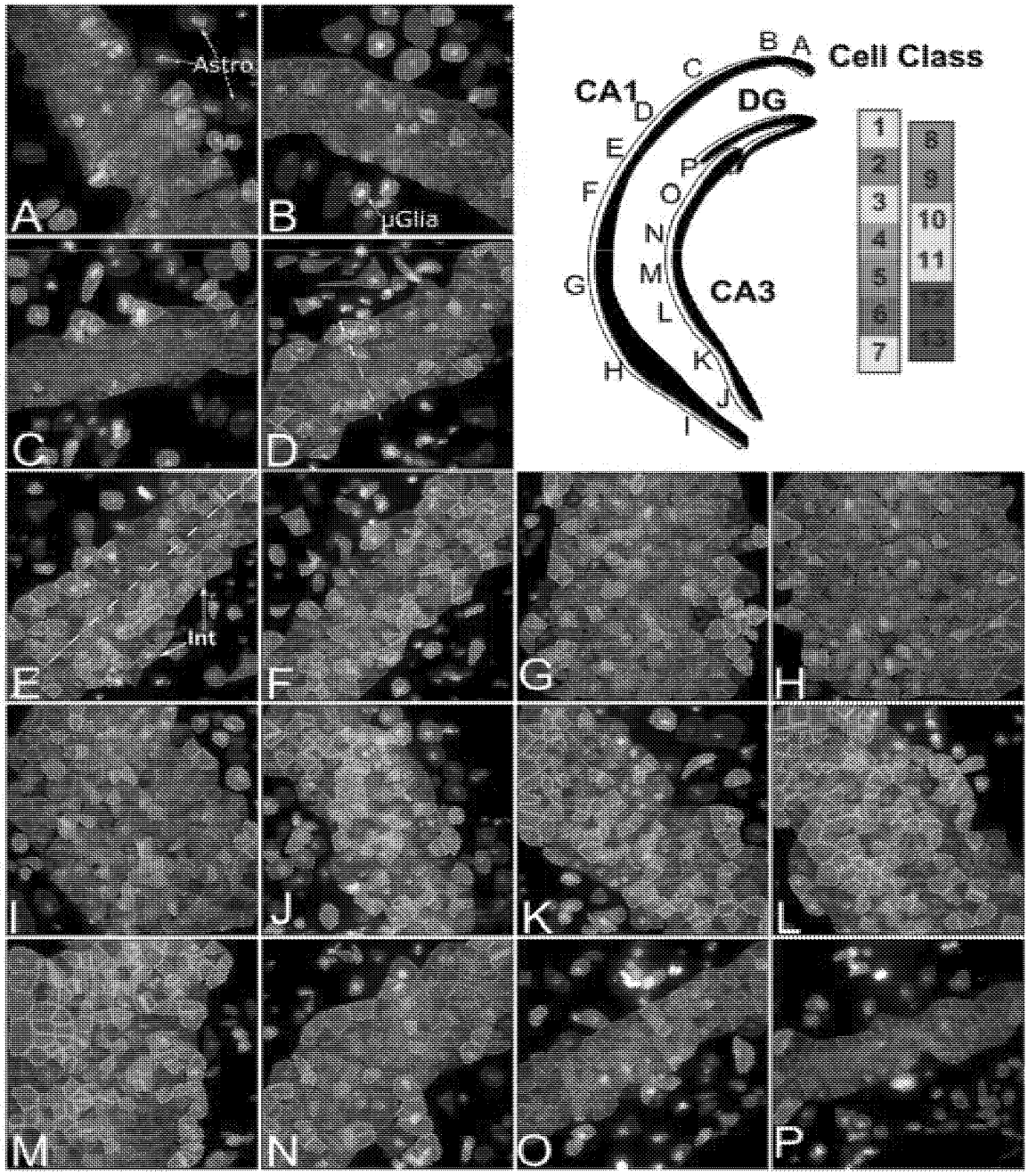

Astro
μGlia
Int
CA1
DG
CA3
Cell Class
1
2
3
4
5
6
7
8
9
10
11
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P

Fig. 36

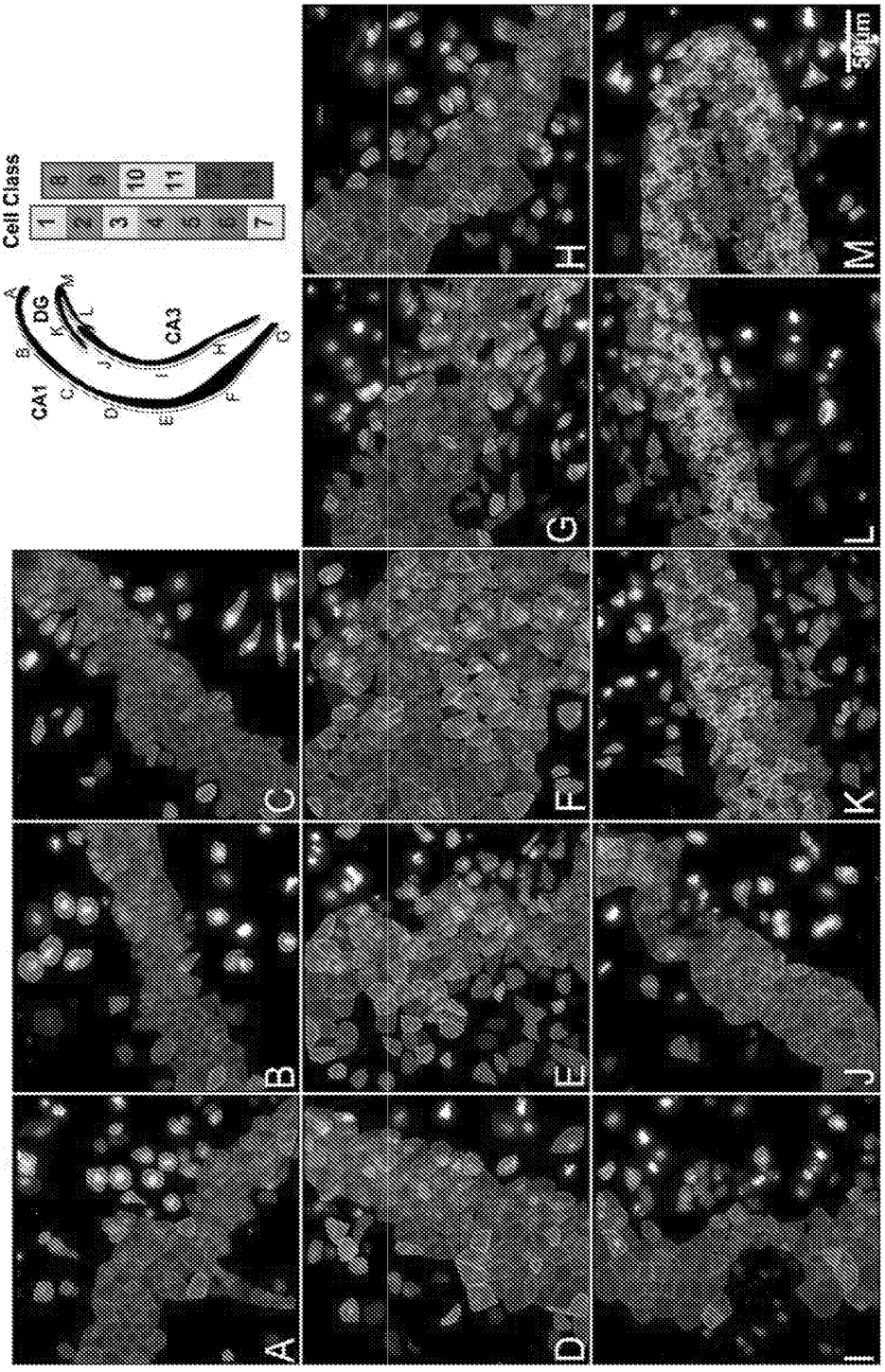


Fig. 37

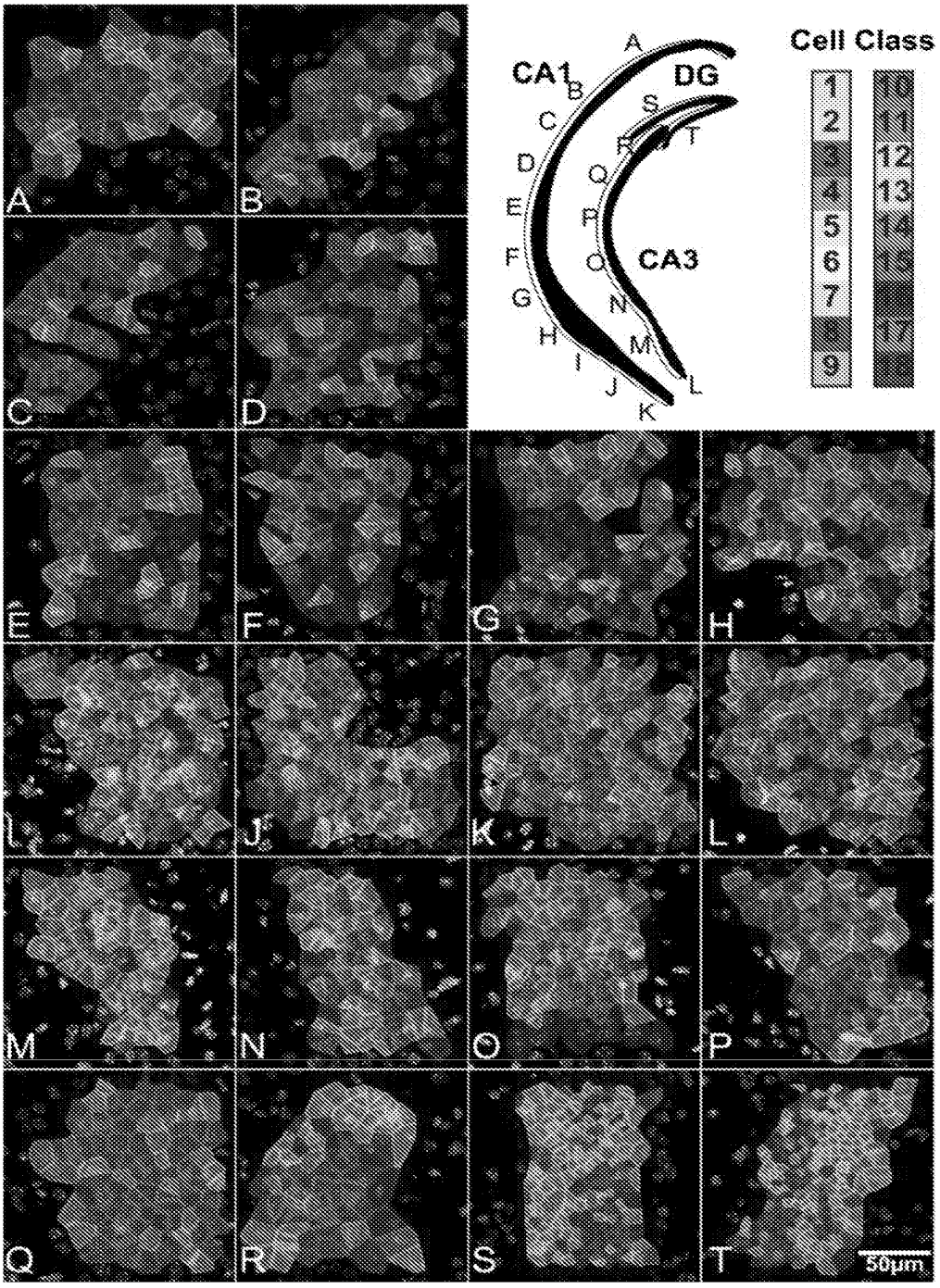

A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
CA1
DG
CA3
Cell Class
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
50µm

Fig. 38

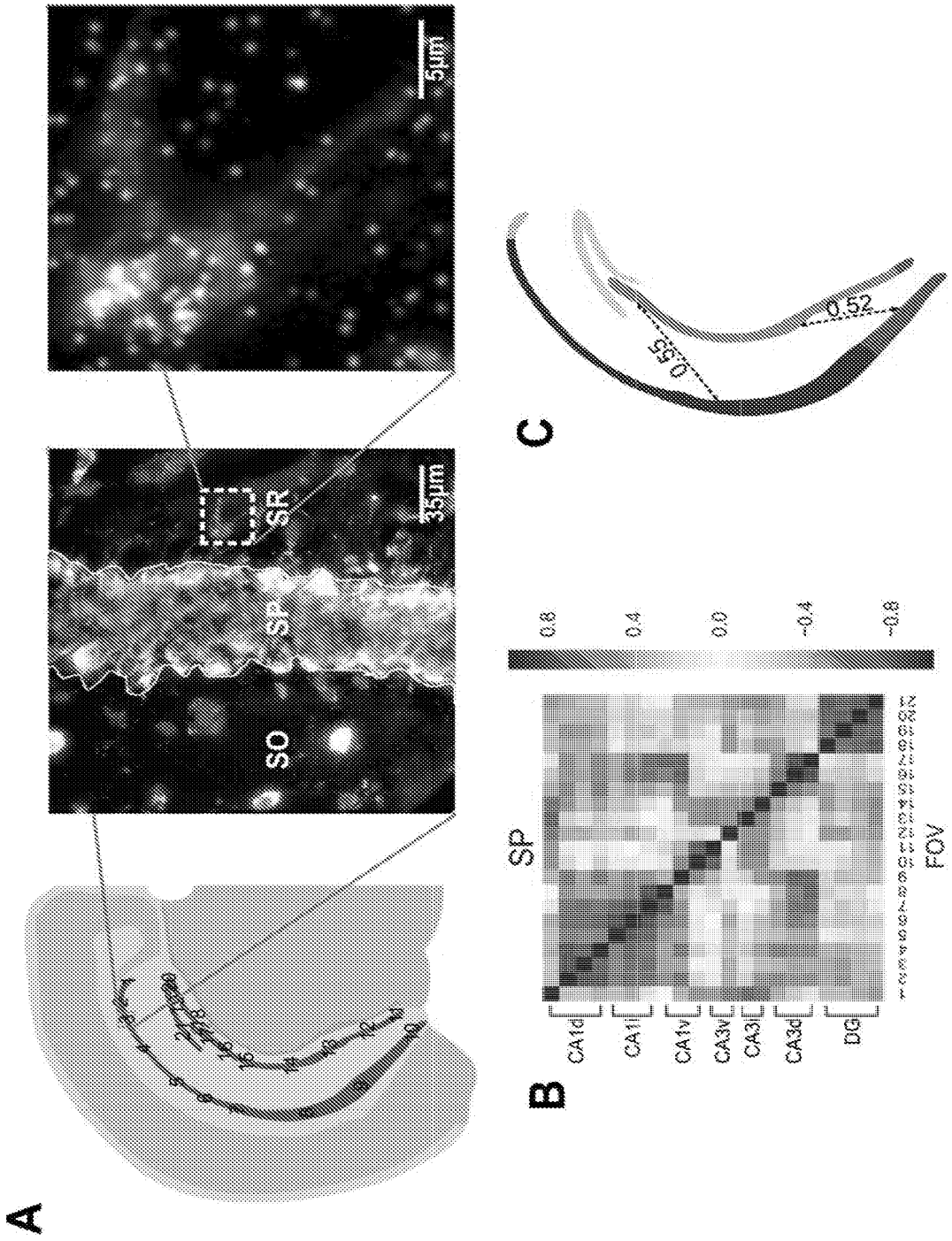

A
5μm
35μm
SR
SP
SO
B
SP
CA1d
CA1i
CA1v
CA3v
CA3i
CA3d
DG
FOV
0.8
0.4
0.0
-0.4
-0.8
C
0.55
0.52

Fig. 39

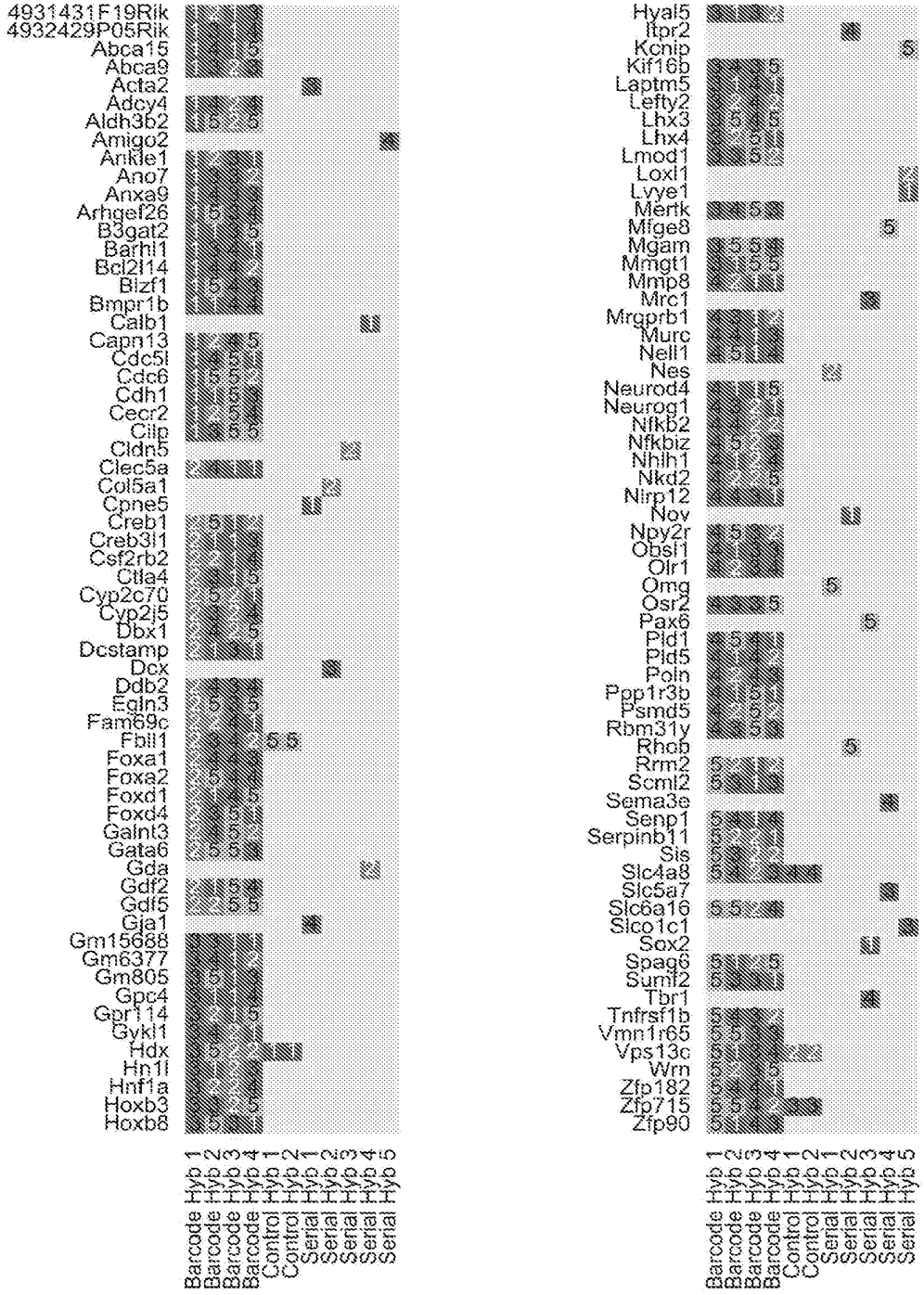

4931431F19Rik
4932429P05Rik
Abca15
Abca9
Acta2
Adcy4
Aldh3b2
Amigo2
Ankle1
Ano7
Anxa9
Arhgef26
B3gat2
Barhl1
Bcl2l14
Blzf1
Bmpr1b
Calb1
Capn13
Cdc5l
Cdc6
Cdh1
Cecr2
Cilp
Cldn5
Clec5a
Col5a1
Cpne5
Creb1
Creb3l1
Csf2rb2
Ctla4
Cyp2c70
Cyp2j5
Dbx1
Dcstamp
Dcx
Ddb2
Egln3
Fam69c
Fbll1
Foxa1
Foxa2
Foxd1
Foxd4
Galnt3
Gata6
Gda
Gdf2
Gdf5
Gja1
Gm15688
Gm6377
Gm805
Gpc4
Gpr114
Gykl1
Hdx
Hn1l
Hnf1a
Hoxb3
Hoxb8
Hyal5
Itpr2
Kcnip
Kif16b
Laptm5
Lefty2
Lhx3
Lhx4
Lmod1
Loxl1
Lvye1
Mertk
Mfge8
Mgam
Mmgt1
Mmp8
Mrc1
Mrgprb1
Murc
Nell1
Nes
Neurod4
Neurog1
Nfkb2
Nfkbiz
Nhlh1
Nkd2
Nlrp12
Nov
Npy2r
Obsl1
Olr1
Omg
Osr2
Pax6
Pld1
Pld5
Poln
Ppp1r3b
Psmd5
Rbm31y
Rhob
Rrm2
Scml2
Sema3e
Senp1
Serpinb11
Sis
Slc4a8
Slc5a7
Slc6a16
Slco1c1
Sox2
Spag6
Sumf2
Tbr1
Tnfrsf1b
Vmn1r65
Vps13c
Wrn
Zfp182
Zfp715
Zfp90
Barcode Hyb 1
Barcode Hyb 2
Barcode Hyb 3
Barcode Hyb 4
Control Hyb 1
Control Hyb 2
Serial Hyb 1
Serial Hyb 2
Serial Hyb 3
Serial Hyb 4
Serial Hyb 5

Fig. 40

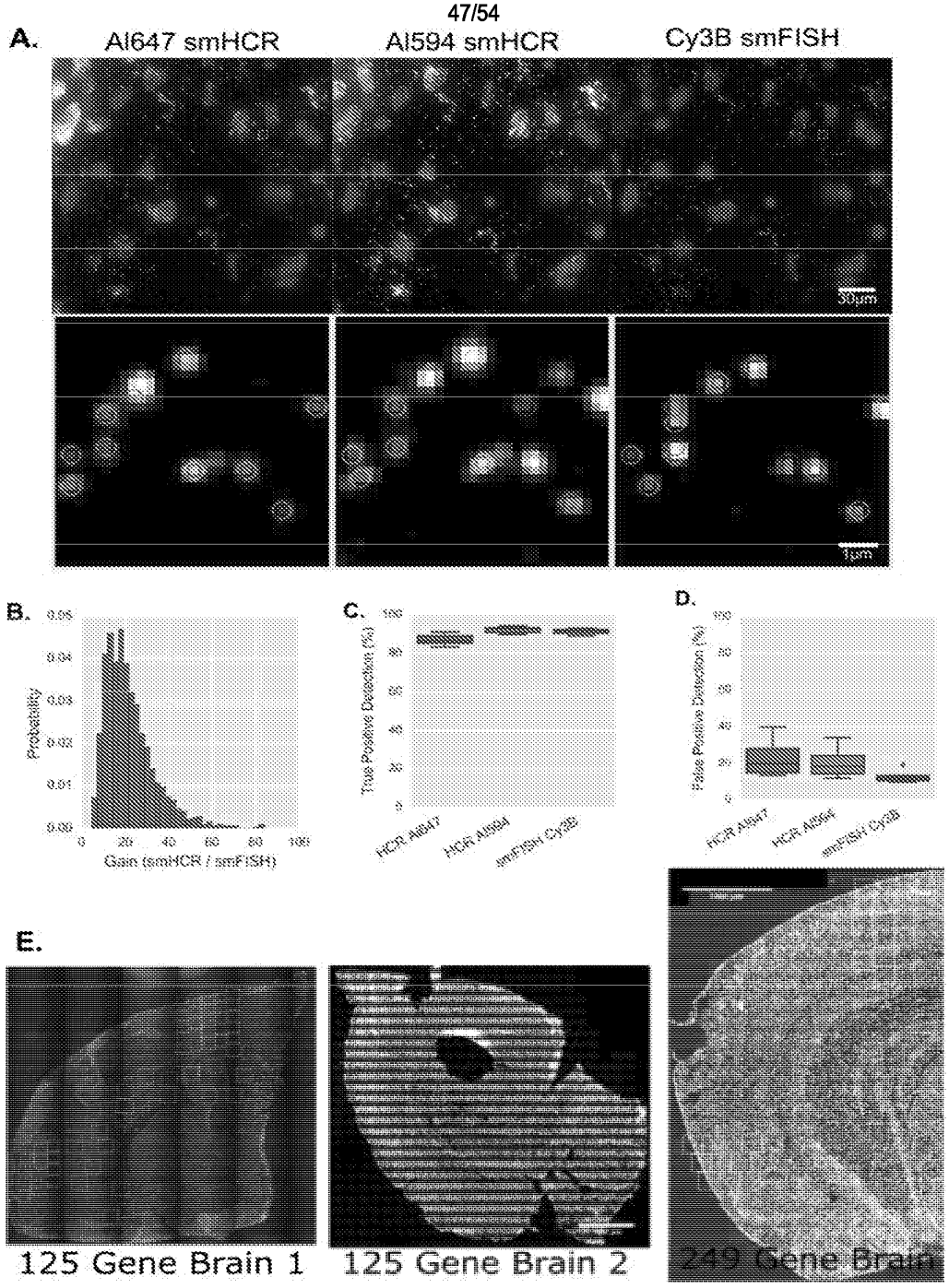

A.
Al647 smHCR
Al594 smHCR
Cy3B smFISH
50μm
1μm
B.
Probability
Gain (smHCR / smFISH)
C.
True Positive Detection (%)
HCR Al647
HCR Al594
smFISH Cy3B
D.
False Positive Detection (%)
HCR Al647
HCR Al594
smFISH Cy3B
E.
125 Gene Brain 1
125 Gene Brain 2
249 Gene Brain

Fig. 41

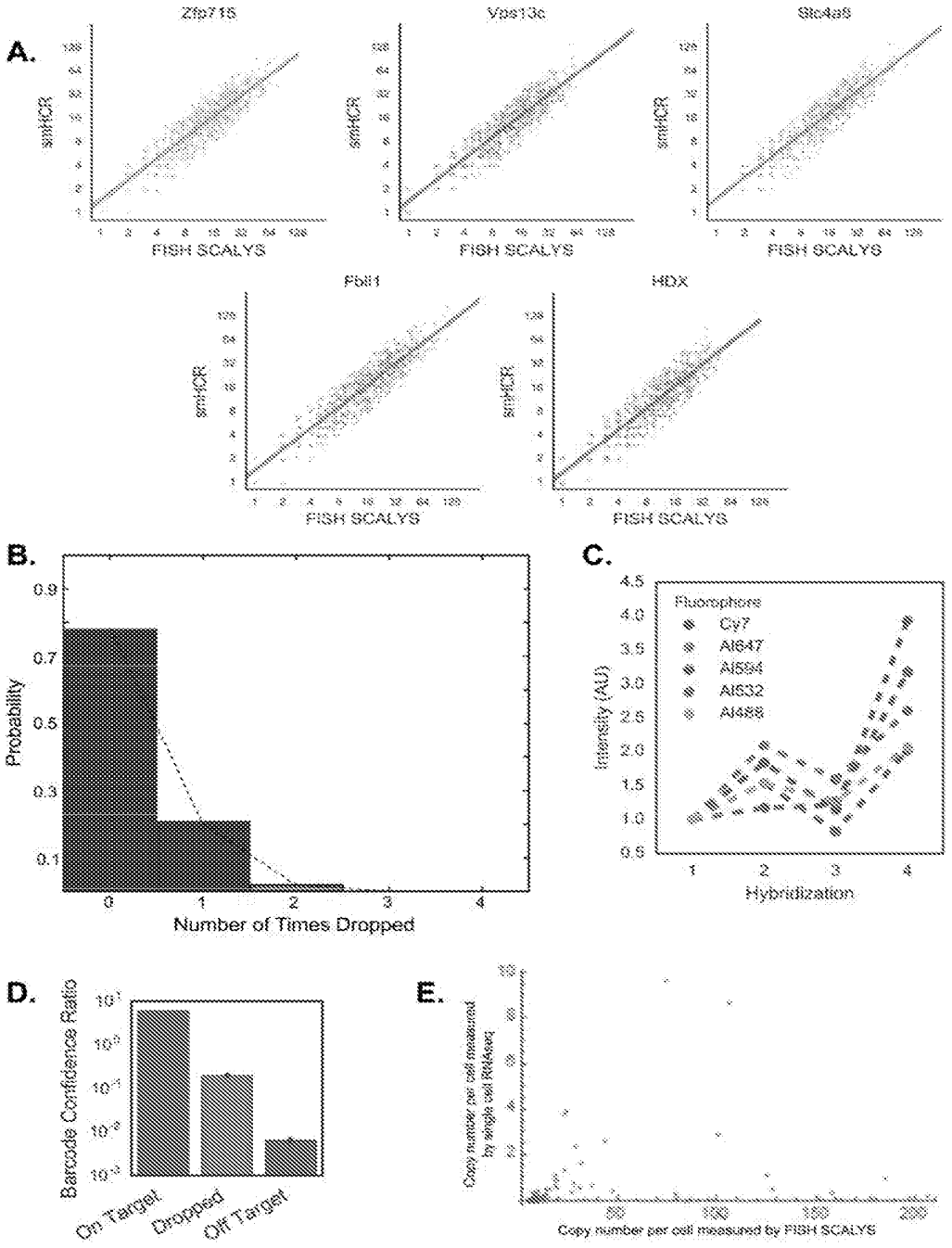

A.
Zfp715
Vps13c
Stc4a8
Pbl1
HDX
smHCR
FISH SCALYS
B.
Probability
Number of Times Dropped
C.
Fluorophore
Cy7
Al647
Al594
Al532
Al488
Intensity (AU)
Hybridization
D.
Barcode Confidence Ratio
On Target
Dropped
Off Target
E.
Copy number per cell measured by single cell RNAseq
Copy number per cell measured by FISH SCALYS

Fig. 42

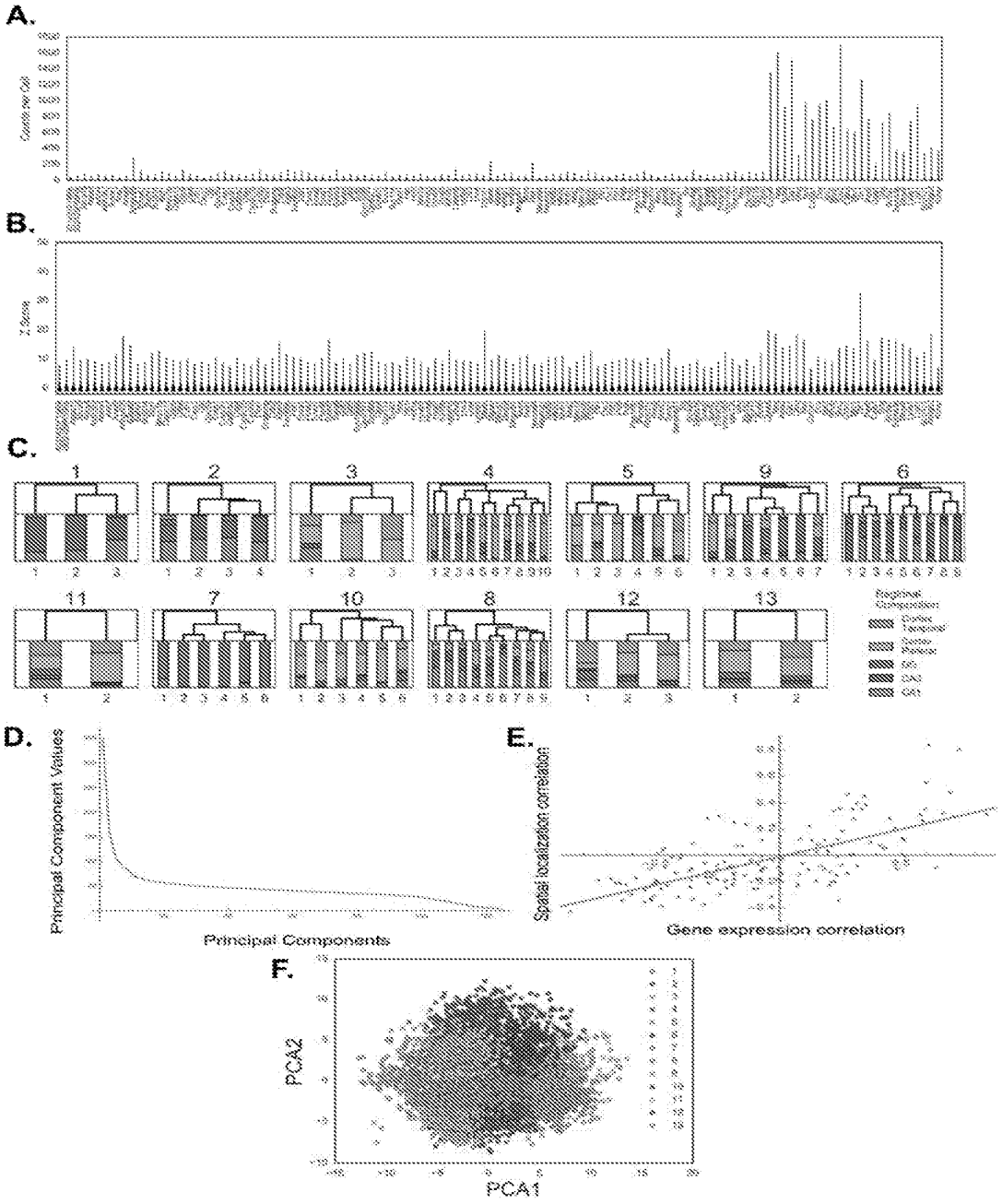

A.
B.
C.
D.
Principal Component Values
Principal Components
E.
Spatial localization correlation
Gene expression correlation
F.
PCA2
PCA1

Fig. 43

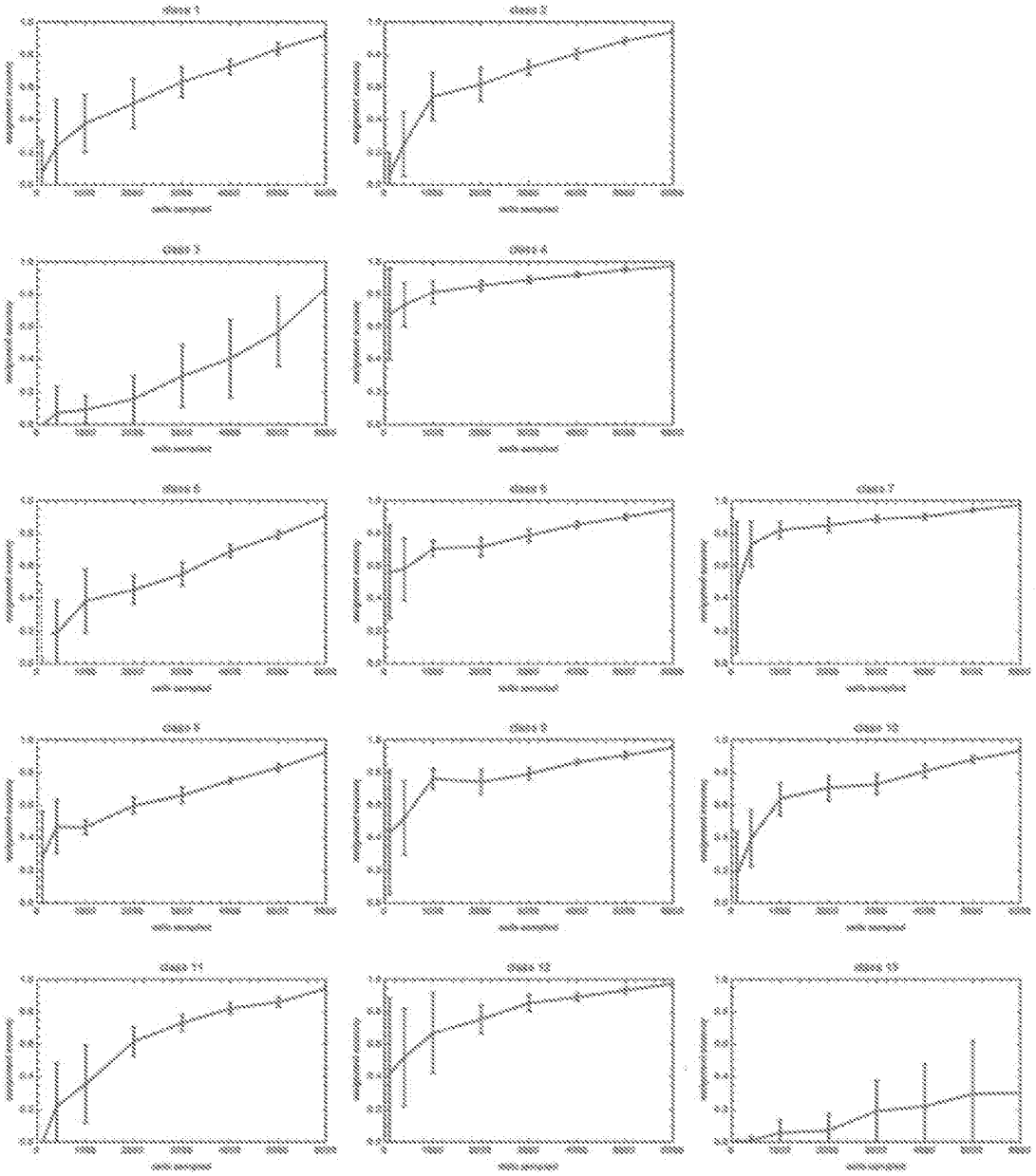


Fig. 44

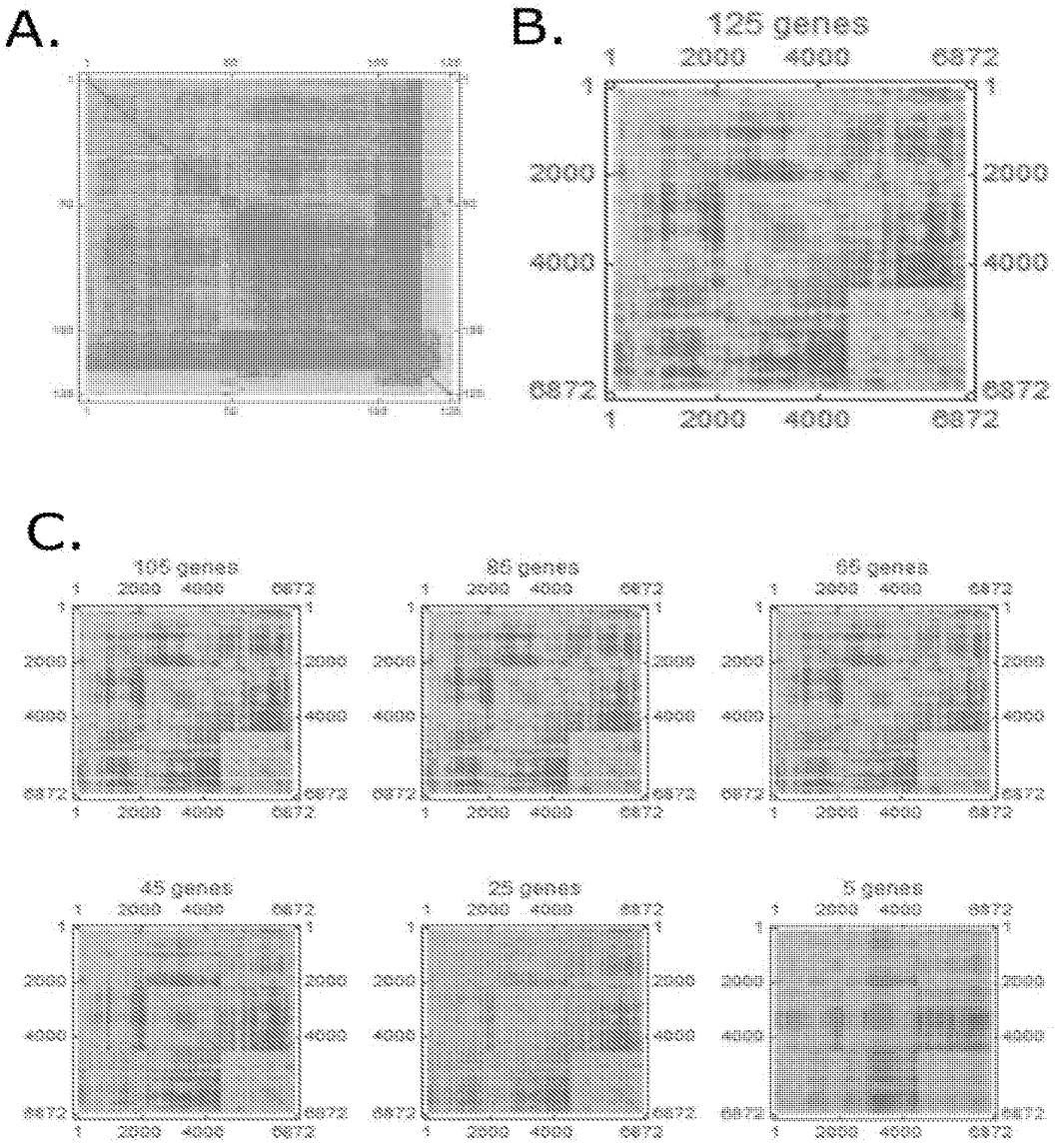

A.
B.
125 genes
1
2000
4000
6872
C.
105 genes
85 genes
65 genes
45 genes
25 genes
5 genes

Fig. 45

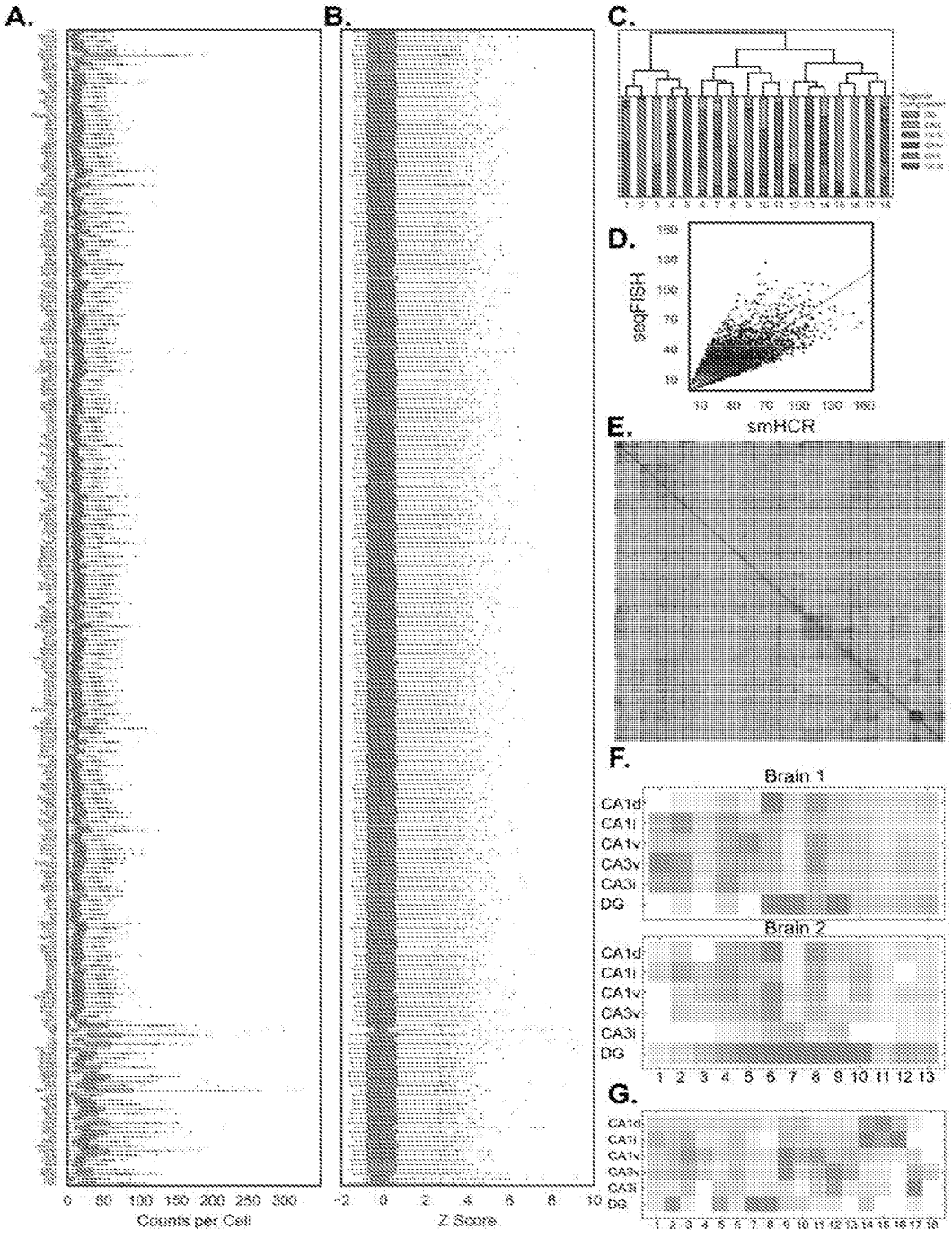

A.
B.
C.
D.
seqFISH
smHCR
E.
F.
Brain 1
Brain 2
CA1d
CA1i
CA1v
CA3v
CA3i
DG
G.
0 50 100 150 200 250 300
Counts per Cell
-2 0 2 4 6 8 10
Z Score

Fig. 46

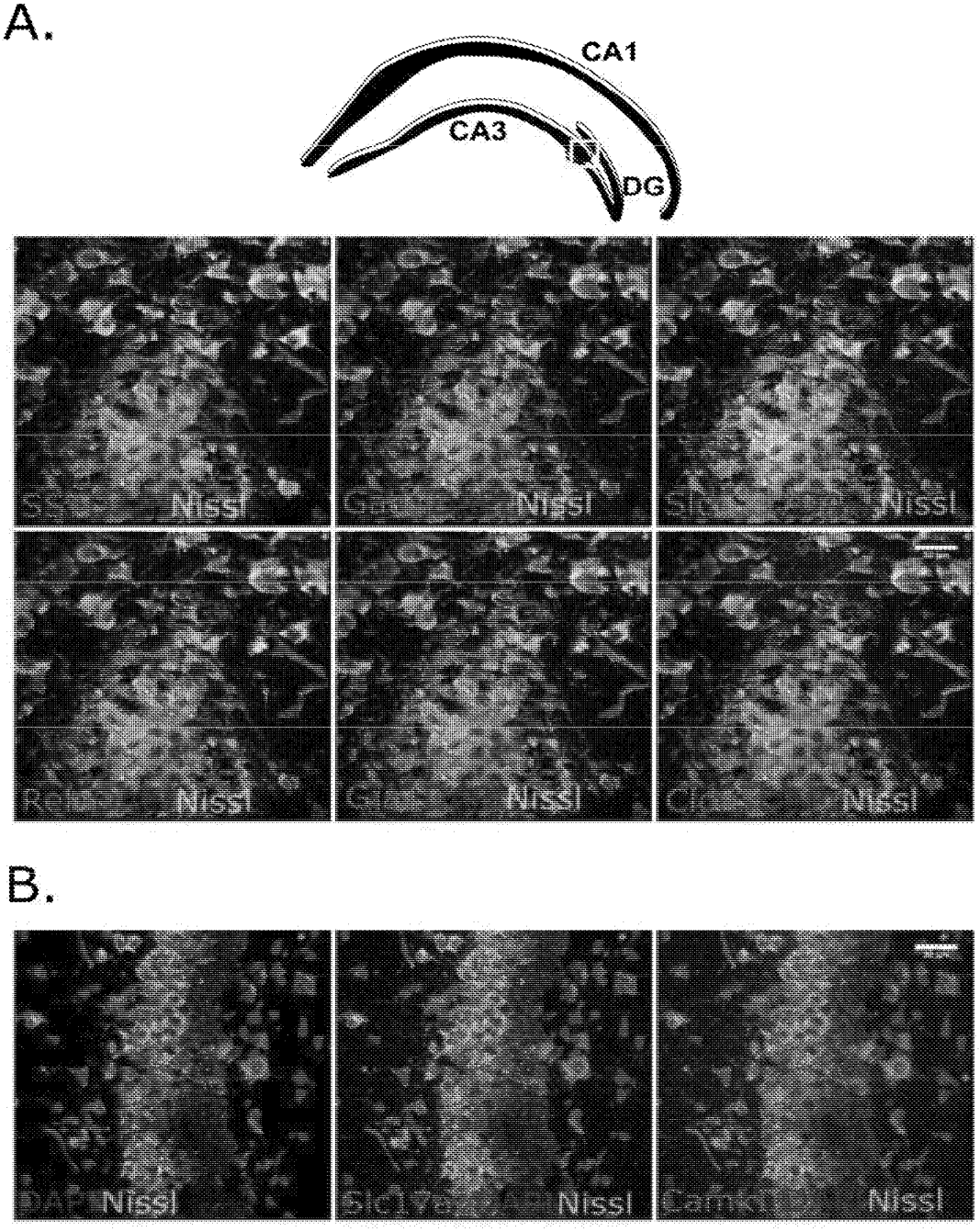

A.
CA1
CA3
DG
Nissl
Nissl
Nissl
Nissl
Nissl
Nissl
B.
DAPI
Nissl
Slc17a7
Nissl
Gamma1
Nissl

Fig. 47

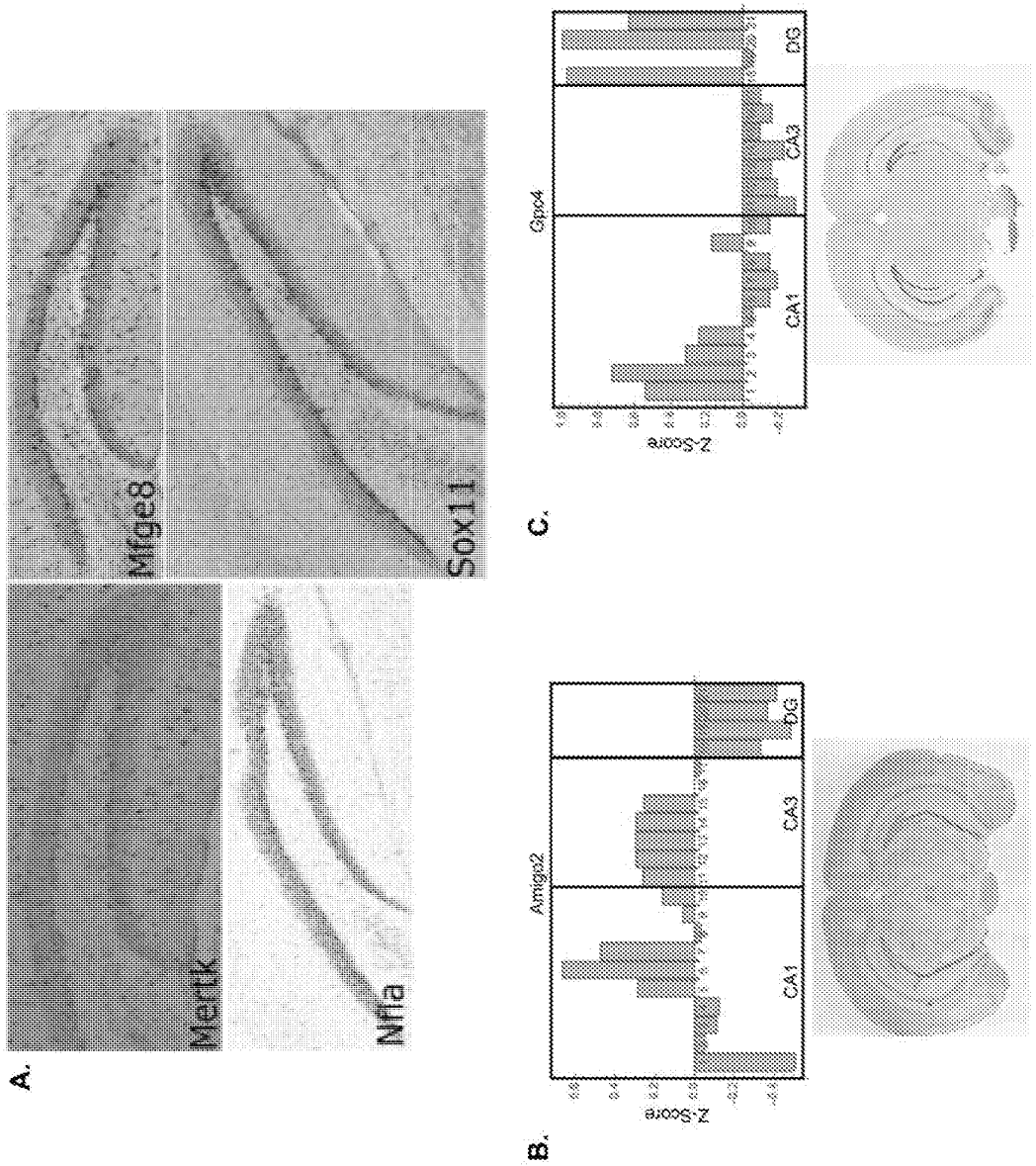

A.
Mfge8
Sox11
Mertk
Nfia
B.
Amigo2
CA1
CA3
DG
Z-Score
C.
Gpx4
CA1
CA3
DG
Z-Score

Fig. 48

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US17/44994

A. CLASSIFICATION OF SUBJECT MATTER
IPC - C12Q 1/68; G01N 21/64 (2017.01)
CPC - G01N 33/5308; G06F 19/24, 19/26; C12Q 1/6813, 1/6841, 1/6804

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	WO 2014/182528 A2 (CALIFORNIA INSTITUTE OF TECHNOLOGY) 13 November 2014; paragraphs abstract, title, paragraphs [0007], [0012], [0031], [0039-40], [0065], [0067], [0069-71], [0087], [0093], [00121], [00161-164], [00168], [00175], [00183-184]; Figures 1-3; Claim 20	1, 2, 3/1, 3/2, 24, 25, 26/24, 26/25, 31 ---------- 30, 32/30, 32/31
Y	WO 2016/018960 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 4 February 2016; page 5, lines 5-9, page 13, lines 8-12, page 25, lines 5-7 and 17-22; page 81, lines 16-19	30, 32/30, 32/31
A	US 2016/0019334 A1 (CALIFORNIA INSTITUTE OF TECHNOLOGY) 21 January 2016; entire document	1, 2, 3/1, 3/2, 24, 25, 26/24, 26/25, 30, 31, 32/30, 32/31
A	US 2010/0323348 A1 (Hamady, ML et al.) 23 December 2010; entire document	1, 2, 3/1, 3/2, 24, 25, 26/24, 26/25, 30, 31, 32/30, 32/31
P, X	(Moffitt, JR, et al.) High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization, PNAS 27 September 2016, volume 113, no. 39, pages 11046–11051 doi: 10.1073/pnas.1612826113; entire document	1, 2, 3/1, 3/2, 24, 25, 26/24, 26/25, 30, 31, 32/30, 32/31

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
29 September 2017 (29.09.2017)	20 OCT 2017
Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US17/44994

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 4-23, 27-29, 33
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (2)) (January 2015)

PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION CONCERNING TRANSMITTAL OF COPY OF INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY (CHAPTER I OF THE PATENT COOPERATION TREATY)
(PCT Rule 44bis.1(c))

To:

CHEN, Tani
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206
ETATS-UNIS D'AMERIQUE

Date of mailing *(day/month/year)* 05 December 2019 (05.12.2019)	
Applicant's or agent's file reference H0498.70637	**IMPORTANT NOTICE**

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US2018/034651	25 May 2018 (25.05.2018)	26 May 2017 (26.05.2017)

Applicant
PRESIDENT AND FELLOWS OF HARVARD COLLEGE

The International Bureau transmits herewith a copy of the international preliminary report on patentability (Chapter I of the Patent Cooperation Treaty).

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No. +41 22 338 82 70

Authorized officer
Yukari Nakamura

e-mail: pct.team7@wipo.int

Form PCT/IB/326 (January 2004)

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY
(Chapter I of the Patent Cooperation Treaty)

(PCT Rule 44*bis*)

Applicant's or agent's file reference H0498.70637	**FOR FURTHER ACTION**	See item 4 below
International application No. PCT/US2018/034651	International filing date *(day/month/year)* 25 May 2018 (25.05.2018)	Priority date *(day/month/year)* 26 May 2017 (26.05.2017)
International Patent Classification (8th edition unless older edition indicated) See relevant information in Form PCT/ISA/237		
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE		

1. This international preliminary report on patentability (Chapter I) is issued by the International Bureau on behalf of the International Searching Authority under Rule 44 *bis*.1(a).

2. This REPORT consists of a total of 9 sheets, including this cover sheet.

   In the attached sheets, any reference to the written opinion of the International Searching Authority should be read as a reference to the international preliminary report on patentability (Chapter I) instead.

3. This report contains indications relating to the following items:

☒	Box No. I	Basis of the report
☐	Box No. II	Priority
☒	Box No. III	Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
☒	Box No. IV	Lack of unity of invention
☒	Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
☐	Box No. VI	Certain documents cited
☐	Box No. VII	Certain defects in the international application
☐	Box No. VIII	Certain observations on the international application

4. The International Bureau will communicate this report to designated Offices in accordance with Rules 44*bis*.3(c) and 93*bis*.1 but not, except where the applicant makes an express request under Article 23(2), before the expiration of 30 months from the priority date (Rule 44*bis* .2).

	Date of issuance of this report 26 November 2019 (26.11.2019)
The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland Facsimile No. +41 22 338 82 70	Authorized officer Yukari Nakamura e-mail: pct.team7@wipo.int

Form PCT/IB/373 (January 2004)

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To: Tani Chen
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
United States of America

PCT

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing *(day/month/year)*	01 OCT 2018
Applicant's or agent's file reference H0498.70637	**FOR FURTHER ACTION** See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US18/34651	25 May 2018 (25.05.2018)	26 May 2017 (26.05.2017)

International Patent Classification (IPC) or both national classification and IPC
IPC - C40B 30/04; C12Q 1/68; G06F 19/20, 19/22, 19/24 (2018.01)
CPC - C12N 15/1065, 15/1079; C12Q 1/68, 1/6813, 1/6841; G06F 19/20, 19/22, 19/24

Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☒ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☒ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☐ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA/US	Date of completion of this opinion	Authorized officer
Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	18 September 2018 (18.09.2018)	Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/237 (cover sheet) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US18/34651

Box No. I Basis of this opinion

1. With regard to the **language**, this opinion has been established on the basis of:
   - ☒ the international application in the language in which it was filed.
   - ☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a)).

3. ☒ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:
   - a. ☒ forming part of the international application as filed:
     - ☒ in the form of an Annex C/ST.25 text file.
     - ☐ on paper or in the form of an image file.
   - b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
   - c. ☐ furnished subsequent to the international filing date for the purposes of international search only:
     - ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).
     - ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US18/34651

Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☒ claims Nos. 5-141, 145-148, 155-158, 165-179, 182-201

because:

☐ the said international application, or the said claims Nos. _____ relate to the following subject matter which does not require an international search *(specify)*:

☒ the description, claims or drawings *(indicate particular elements below)* or said claims Nos. -***-Please See Below-***- are so unclear that no meaningful opinion could be formed *(specify)*:

Claims Nos.: 5-141, 145-148, 155-158, 165-179, 182-201 are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

☐ the claims, or said claims Nos. _____ are so inadequately supported by the description that no meaningful opinion could be formed *(specify)*:

☒ no international search report has been established for said claims Nos. 5-141, 145-148, 155-158, 165-179, 182-201

☐ a meaningful opinion could not be formed without the sequence listing; the applicant did not, within the prescribed time limit:

- ☐ furnish a sequence listing in the form of an Annex C/ST.25 text file, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.
- ☐ furnish a sequence listing on paper or in the form of an image file complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.
- ☐ pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rule 13*ter*.1(a) or (b).

☐ See Supplemental Box for further details.

Form PCT/ISA/237 (Box No. III) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US18/34651

Box No. IV Lack of unity of invention

1. ☒ In response to the invitation (Form PCT/ISA/206) to pay additional fees the applicant has, within the applicable time limit:

   ☐ paid additional fees.

   ☐ paid additional fees under protest and, where applicable, the protest fee.

   ☐ paid additional fees under protest but the applicable protest fee was not paid.

   ☒ not paid additional fees.

2. ☐ This Authority found that the requirement of unity of invention is not complied with and chose not to invite the applicant to pay additional fees.

3. This Authority considers that the requirement of unity of invention in accordance with Rule 13.1, 13.2 and 13.3 is

   ☐ complied with.

   ☒ not complied with for the following reasons:

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, Claims 1-4, 142-144, 149-154, 159, 161-164, 180-181 are directed toward introducing, into a plurality of cells, nucleic acids comprising an identification portion and an expression portion and imaging cells, for example, to determine phenotypes and/or genotypes in populations of cells.

Group II, Claim 160 is directed toward a YFAST protein.

The inventions listed as Groups I and II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical features of Group I include nucleic acid tags and probes, not present in Group II; the special technical features of Group II include SEQ ID NO: 3, not present in Group I.

Groups I and II share the technical features including: a detection tag.

However, these shared technical features are previously disclosed by WO 2016/018960 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) (hereinafter 'Harvard').

Harvard discloses a detection tag (page 46, lines 4-10; figure 1B).

Since none of the special technical features of the Groups I and II inventions are found in more than one of the inventions, and since all of the shared technical features are previously disclosed by the Harvard reference, unity of invention is lacking.

4. Consequently, this opinion has been established in respect of the following parts of the international application:

   ☐ all parts.

   ☒ the parts relating to claims Nos. -***-Please See Above-***-

Form PCT/ISA/237 (Box No. IV) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US18/34651

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	-***-Please See Below-***-	YES
	Claims	150, 161-162, 180-181	NO
Inventive step (IS)	Claims	NONE	YES
	Claims	-***-Please See Below-***-	NO
Industrial applicability (IA)	Claims	-***-Please See Below-***-	YES
	Claims	NONE	NO

2. Citations and explanations:

-Continued from Box No. V: 1. Statement:

Novelty (N): YES: Claims 1-3, 4/2-3, 142-143, 144/142-143, 149, 151-153, 154/152-153, 159, 163, 164/162-163, 180-181

Inventive step (IS): NO: Claims 1-3, 4/2-3, 142-143, 144/142-143, 149-153, 154/152-153, 159, 161-163, 164/162-163, 180-181

Industrial Applicability (IA): YES: Claims 1-3, 4/2-3, 142-143, 144/142-143, 149-153, 154/152-153, 159, 161-163, 164/162-163, 180-181

-Continued from Box No. V. 2: Citations and Explanations:

Claim 150 lacks novelty under PCT Article 33(2) as being anticipated by WO 1999/031277 A1 (Medical Science Systems, Inc.) (hereinafter 'MSI').
As per claim 150, MSI discloses a method (abstract), comprising: acquiring an image of a plurality of cells transfected with a plurality of nucleic acids comprising an identification portion and an expression portion (performing confocal microscopy on a plurality of cells transfected with a test promoter coupled with a reporter; abstract; page 14, lines 15-25), wherein the cells exhibit imagable differences in phenotype due to expression of the expression portion (wherein cells fluoresce (wherein the cells exhibit imagable differences in phenotype due to expression of the expression portion); page 14, line 15 – page 15, line 7); and acquiring a plurality of images of the plurality of cells (page 14, lines 15-25; figure 2), wherein the images of the cells exhibit differences due to differences in the identification portions of the nucleic acids within the cells (wherein the cells exhibit imagable differences in phenotype due to expression of the expression portion); page 14, line 15 – page 15, line 7).

Claims 161-162 and 180-181 lack novelty under PCT Article 33(2) as being anticipated by US 2006/0051798 A1 to Mirkin et al. (hereinafter 'Mirkin').

As per claim 161, Mirkin discloses a method (abstract; claim 1), comprising: creating plurality of nucleic acids comprising an identification portion and an expression portion (creating plurality of nucleic acids comprising a DNA barcode (an identification portion) and a specific binding complement to a target analyte (an expression portion); paragraph [0162]; claim 1), wherein when expressed in cells, the identification portion is identifiable by sequentially exposing the cells to nucleic acid probes (wherein when the target analyte is within a cell (expressed in cells), the identification portion is identifiable by sequentially exposing the cells to nucleic acid probes; paragraph [0087]; claim 1).

As per claim 162, Mirkin discloses the method of claim 161, and Mirkin further discloses pairing an identification portion and an expression portion to form a nucleic acid (the nucleic acid comprises a DNA barcode (an identification portion) having some nucleic acids complementary to a specific binding complement to a target analyte (an expression portion); claim 1).

As per claim 180, Mirkin discloses a composition (paragraph [0004]; claim 1), comprising: a plurality of nucleic acids comprising an identification portion and an expression portion (comprising a plurality of nucleic acids comprising a DNA barcode (an identification portion) and a specific binding complement to a target analyte (an expression portion); paragraph [0132]; claim 1), wherein when expressed in cells, the identification portion is identifiable by sequentially exposing the cells to nucleic acid probes (wherein when the target analyte is within a cell (expressed in cells), the identification portion is identifiable by sequentially exposing the cells to nucleic acid probes; paragraph [0087]; claim 1).

As per claim 181, Mirkin discloses the composition of claim 180, and Mirkin further discloses wherein the identification portion of the plurality of nucleic acids comprises at least 10 unique sequences (wherein the identification portion of the barcode comprises 4^12 different sequences (at least 10 unique sequences); paragraph [0112]).

-***-Continued Within the Next Supplemental Box-***-

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US18/34651

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

-***-Continued from Box No. V: Citations and Explanations: -***-

Claims 1-2, and 4/2 lack an inventive step under PCT Article 33(3) as being obvious over US 6,524,798 B1 to Goldbard et al. (hereinafter 'Goldbard') in view of MSI.

As per claim 1, Goldbard discloses a method (abstract; claim 1), comprising: introducing, into a plurality of cells, antibodies comprising an identification portion (introducing into a plurality of cells an antibody bound to a ligand, wherein the antibody is specific for an antigen suspected of being associated with the cell (an identification portion); abstract; column 8, lines 33-35; claim 1); determining phenotype of the cells (abstract; claim 1); and determining genotype of the plurality of cells by sequentially exposing the plurality of cells to nucleic acid probes (determining the genotype of the cells by sequentially contacting the cells with polynucleotide probes; claim 1), and determining binding of target sequences of the nucleic acid probes within the plurality of cells (detecting the presence of the complex of the polynucleotide probe and cellular nucleic acid, and correlating the presence of said complex of the polynucleotide probe and cellular nucleic acid with the presence of the target sequence in the cell, thereby determining the cell genotype; claim 1). Goldbard does not disclose introducing nucleic acids comprising an identification portion and an expression portion. MSI discloses a method for detecting a phenotype (abstract), comprising introducing, into a plurality of cells, nucleic acids comprising an identification portion and an expression portion (transfecting an appropriate eukaryotic host cell, a test DNA sequence comprising a test promoter coupled with a reporter; abstract). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Goldbard, to include introducing nucleic acids comprising an identification portion and an expression portion, as previously disclosed by MSI, in order to provide a superior nucleic acid that can be identified based on its identification portion.

As per claim 2, Goldbard and MSI, in combination, disclose the method of claim 1, but Goldbard does not disclose wherein introducing the nucleic acid comprises transfecting the nucleic acid into the plurality of cells. MSI discloses wherein introducing the nucleic acid comprises transfecting the nucleic acid into the plurality of cells (transfecting an appropriate eukaryotic host cell, a test DNA sequence comprising a test promoter coupled with a reporter; abstract). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Goldbard, to include wherein introducing the nucleic acid comprises transfecting the nucleic acid into the plurality of cells, as previously disclosed by MSI, in order to provide a superior nucleic acid that can be identified based on its identification portion.

As per claim 4/2, Goldbard and MSI, in combination, disclose the method of claim 2, and Goldbard further discloses wherein the transfected cells are sorted (labeled cells are detected by FACS analysis; column 11, lines 33-34). Goldbard does not disclose transfected cells. MSI discloses transfected cells (abstract). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Goldbard, to include transfected cells, as previously disclosed by MSI, in order to provide a superior cell that has been transfected for easy identification.

Claims 142-143, 144/142-143, 151-153, 154/152-153, and 159 lack an inventive step under PCT Article 33(3) as being obvious over MSI in view of WO 2016/018960 A1 (President And Fellows Of Harvard College) (hereinafter 'Harvard').

As per claim 142, MSI discloses a method (abstract), comprising: introducing, into a plurality of cells, a nucleic acid comprising an identification portion and an expression portion (transfecting an appropriate eukaryotic host cell, a test DNA sequence comprising a test promoter coupled with a reporter; abstract); imaging the plurality of cells to determine phenotype of the cells (determining the phenotype by FACS; page 10, lines 13-18). MSI does not disclose and determining genotype of the plurality of cells. Harvard discloses a method (abstract), comprising introducing, into a plurality of cells, a nucleic acid (a plurality of nucleic acid probes may be applied to a cell sample, and their binding within the sample determined using fluorescence, to determine locations of the nucleic acid probes within the sample; abstract) and determining genotype of the plurality of cells (determining the entire genome of the cell; page 17, lines 18-20). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of MSI, to include genotype of the plurality of cells, as previously disclosed by Harvard, in order to provide an improved method for identifying a cell based on phenotype and genotype.

As per claim 143, MSI and Harvard, in combination, disclose the method of claim 142, but MSI does not disclose wherein determining genotype comprises using MERFISH. Harvard discloses wherein determining genotype comprises using MERFISH (page 17, lines 18-20; page 65, lines 1-3). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of MSI, to include wherein determining genotype comprises using MERFISH, as previously disclosed by Harvard, in order to provide an improved method for determining the genotype of a cell.

As per claims 144/142-143, MSI and Harvard, in combination, disclose the method of claims 142-143, but MSI does not disclose wherein determining genotype comprises using smFISH. Harvard discloses wherein determining genotype comprises using smFISH (gene expression is determined by MERFISH, which is a highly multiplexed smFISH imaging method; page 17, lines 18-20; page 56, lines 8-10; page 65, lines 1-3). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of MSI, to include wherein determining genotype comprises using smFISH, as previously disclosed by Harvard, in order to provide an improved method for determining the genotype of a cell.

-***-Continued Within the Next Supplemental Box-***-

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US18/34651

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

-***-Continued from Previous Supplemental Box-***-

As per claim 151, MSI discloses a method (abstract) comprising: acquiring an image of a plurality of cells transfected with a plurality of nucleic acids comprising an identification portion and an expression portion (performing confocal microscopy on a plurality of cells transfected with a test promoter coupled with a reporter; abstract; page 14, lines 15-25), wherein the cells exhibit imagable differences in phenotype due to expression of the expression portion (wherein cells fluoresce (wherein the cells exhibit imagable differences in phenotype due to expression of the expression portion); page 14, line 15 – page 15, line 7); and performing FACS (page 17, lines 4-6). MSI does not disclose and performing in situ hybridization, FISH, multiplexed FISH, smFISH, CASFISH, and/or MERFISH on the cells. Harvard discloses performing MERFISH (page 17, lines 18-20; page 65, lines 1-3). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of MSI, to include performing MERFISH, as previously disclosed by Harvard, in order to provide an improved method for determining the phenotype of a cell.

As per claim 152, MSI discloses a method (abstract), comprising: acquiring an image of a plurality of cells transfected with a plurality of nucleic acids comprising an identification portion and an expression portion (performing FACS on a plurality of cells transfected with a test promoter coupled with a reporter; abstract; page 10, lines 13-18), and determining genotype of the plurality of cells. MSI does not disclose and determining genotype of the plurality of cells. Harvard discloses a method (abstract), comprising introducing, into a plurality of cells, a nucleic acid (a plurality of nucleic acid probes may be applied to a cell sample, and their binding within the sample determined using fluorescence, to determine locations of the nucleic acid probes within the sample; abstract) and determining genotype of the plurality of cells (determining the entire genome of the cell; page 17, lines 18-20). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of MSI, to include genotype of the plurality of cells, as previously disclosed by Harvard, in order to provide an improved method for identifying a cell based on genotype.

As per claim 153, MSI and Harvard, in combination, disclose the method of claim 152, but MSI does not disclose wherein determining genotype comprises using MERFISH. Harvard discloses wherein determining genotype comprises using MERFISH (page 17, lines 18-20; page 65, lines 1-3). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of MSI, to include wherein determining genotype comprises using MERFISH, as previously disclosed by Harvard, in order to provide an improved method for determining the genotype of a cell.

As per claims 154/152-153, MSI and Harvard, in combination, disclose the method of claims 152-153, but MSI does not disclose wherein determining genotype comprises using smFISH. Harvard discloses wherein determining genotype comprises using smFISH (gene expression is determined by MERFISH, which is a highly multiplexed smFISH imaging method; page 17, lines 18-20; page 56, lines 8-10; page 65, lines 1-3). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of MSI, to include wherein determining genotype comprises using smFISH, as previously disclosed by Harvard, in order to provide an improved method for determining the genotype of a cell.

As per claim 159, MSI discloses a method (abstract), comprising: introducing, into a plurality of cells, a nucleic acid comprising an identification portion and an expression portion (transfecting an appropriate eukaryotic host cell, a test DNA sequence comprising a test promoter coupled with a reporter; abstract); and determining phenotype of the plurality of cells (determining the phenotype by FACS; page 10, lines 13-18). MSI does not disclose by sequentially exposing the plurality of cells to nucleic acid probes, and determining binding of the target sequences of the nucleic acid probes within the plurality of cells. Harvard discloses sequentially exposing the plurality of cells to nucleic acid probes (sequentially applying a plurality of nucleic acid probes to a cell sample; abstract; page 14, lines 22-26), and determining binding of the target sequences of the nucleic acid probes within the plurality of cells (the binding of the probes with the target sequences within the sample are determined using fluorescence, which identifies the locations of the nucleic acid probes within the sample; abstract). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of MSI, to include by sequentially exposing the plurality of cells to nucleic acid probes, and determining binding of the target sequences of the nucleic acid probes within the plurality of cells, as previously disclosed by Harvard, in order to provide an improved method for identifying a cell based on phenotype.

Claim 149 lacks an inventive step under PCT Article 33(3) as being obvious over Mirkin in view of US 2008/0269068 A1 to Church et al. (hereinafter 'Church').

As per claim 149, Mirkin discloses a composition (paragraph [0004]; claim 1), comprising: a plurality of nucleic acids (a plurality of nucleic acids; paragraph [0125]; claim 1) comprising an identification portion and an expression portion (comprising a DNA barcode (an identification portion) and a specific binding complement to a target analyte (an expression portion); claim 1), wherein the identification portion comprises N variable portions, N being at least 2 (wherein the DNA barcode has a sequence having at least two portions; claim 1), each variable portion being of at least two possibilities (the DNA barcode has sequences that are different and that serve as an identifier for a particular target analyte; claim 1), and wherein each possible combination of variable portions is present within the plurality of nucleic acids (each type of particle complex probe contains a barcode for a particular target analyte; paragraph [0115]). Mirkin does not disclose wherein the identification portion comprises N variable portions, N being at least 3. Church discloses a barcode comprising an identification portion (paragraph [0063]), wherein the identification portion comprises N variable portions, N being at least 3 (each barcode has three main parts: 1) a degenerate 'query' portion which base pairs with the unknown region of the template; 2) a fixed portion which directs the degenerate 'query' portion to the unknown region by base pairing with fixed sequence on each side of the unknown tag, and 3) barcode which correlates with the identity of one base in the degenerate 'query' portion and is interrogated by hybridization once on the instrument; paragraph [0063]). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Mirkin, to include wherein the identification portion comprises N variable portions, N being at least 3, as previously disclosed by Church, in order to provide a superior barcode for use in the composition.

-***-Continued Within the Next Supplemental Box-***-

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No. PCT/US18/34651

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**
Continuation of:

-***-Continued from Previous Supplemental Box-***-

Claims 163 and 164/162-163 lack an inventive step under PCT Article 33(3) as being obvious over Mirkin in view of MSI.

As per claim 163, Mirkin discloses the method of claim 162, but Mirkin does not disclose randomly pairing an identification portion and an expression portion to form a nucleic acid. MSI discloses randomly pairing an identification portion and an expression portion to form a nucleic acid (fusing random promoters to GFP proteins for FACS identification; page 10, lines 19-21). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Mirkin, to include randomly pairing an identification portion and an expression portion to form a nucleic acid, as previously disclosed by MSI, in order to provide an improved library of nucleic acids that exhibits random identification portions and an expression portions.

As per claim 164/162, Mirkin discloses the method of claim 162, but Mirkin does not disclose determining pairing between expression portion and identification. MSI discloses determining pairing between expression portion and identification (identifying the test promoter and reporter combinations; page 23, lines 12-16). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Mirkin, to include determining pairing between expression portion and identification, as previously disclosed by MSI, in order to provide an improved library of nucleic acids that exhibits random identification portions and an expression portions.

As per claim 164/163, Mirkin and MSI, in combination, disclose the method of claim 163, but Mirkin does not disclose determining pairing between expression portion and identification. MSI discloses determining pairing between expression portion and identification (identifying the test promoter and reporter combinations; page 23, lines 12-16). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Mirkin, to include determining pairing between expression portion and identification, as previously disclosed by MSI, in order to provide an improved library of nucleic acids that exhibits random identification portions and an expression portions.

Claims 3 and 4/3 lack an inventive step under PCT Article 33(3) as being obvious over Goldbard in view of MSI, and in further view of US 2004/0137488 A1 to Kenten et al. (hereinafter 'Kenten').

As per claim 3, Goldbard and MSI, in combination, disclose the method of claim 2, but Goldbard does not disclose wherein the transfected cells are selected using antibiotic selection. Kenten discloses a method of transfecting a cell with a gene of interest and detecting changes in the phenotype of the cell (paragraph [0018]), wherein the transfected cells are selected using antibiotic selection (wherein the transfected cells are selected using antibiotic selection; paragraph [0037])

As per claim 4/3, Goldbard, MSI, and Kenten, in combination, disclose the method of claim 3, and Goldbard further discloses wherein the transfected cells are sorted (labeled cells are detected by FACS analysis; column 11, lines 33-34). Goldbard does not disclose transfected cells. MSI discloses transfected cells (abstract). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Goldbard, to include transfected cells, as previously disclosed by MSI, in order to provide a superior cell that has been transfected for easy identification.

Claims 1-3, 4/2-3, 142-143, 144/142-143, 149-153, 154/152-153, 159, 161-163, 164/162-163, and 180-181 have industrial applicability as defined by PCT Article 33(4) because the subject matter can be made or used in industry.

Wolf, Greenfield & Sacks, P.C.
IFD MAR 06 2020
Docketed ✓ Already Docketed
Not Required
Initials 1st 2nd

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

PCT

To: Tani Chen
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
United States of America

INVITATION TO PAY ADDITIONAL FEES
AND, WHERE APPLICABLE, PROTEST FEE

(PCT Article 17(3)(a) and Rules 40.1 and 40.2(e))

Date of mailing *(day/month/year)*	03 MAR 2020
Applicant's or agent's file reference: H0498. 70680WO00	**PAYMENT DUE** within **ONE MONTH** from the above date of mailing
International application No.: PCT/US19/65857	International filing date *(day/month/year)*: 12 December 2019 (12.12.2019)
Applicant: PRESIDENT AND FELLOWS OF HARVARD COLLEGE	

1. This International Searching Authority

   (i) considers that there are 3 *(number of)* inventions claimed in the international application covered by the claims indicated below/on an extra sheet:

   -***-Please See Supplemental Page-***-

   (ii) therefore considers that **the international application does not comply with the requirement of unity of invention** (Rules 13.1, 13.2 and 13.3) for the reasons indicated below/on an extra sheet:

   -***-Please See Supplemental Page-***-

   (iii) ☐ has carried out a partial international search (see Annex) ☒ will establish the international search report on those parts of the international application which relate to the invention first mentioned in claims Nos.:

   Group I, Claims 1-3, 100-114, 116-120, 122-133, 139-145, 148-152, 155-159, 162-166, 170-174, and 178-183

   (iv) will establish the international search report on the other parts of the international application only if, and to the extent to which, additional fees are paid.

2. Consequently, the applicant is hereby **invited to pay**, within the time limit indicated above, additional fees in the amount indicated below:

   $2,080.00 (Fee per additional invention) x 2 (number of additional inventions) = $4,160.00 (See Item 2 of next sheet) (total amount of additional fees/currency)

3. The applicant is informed that, according to Rule 40.2(c), **the payment of any additional fees may be made under protest**, that is, a reasoned statement to the effect that the international application complies with the requirement of unity of invention or that the amount of the required additional fees is excessive, where applicable, subject to the payment of a protest fee.

   Where the applicant pays additional fees under protest, the applicant is hereby invited, within the time limit indicated above, to pay a protest fee (Rule 40.2(e)) in the amount of _____ *(amount/currency)*

   Where the applicant has not, within the time limit indicated above, paid the required protest fee, the protest will be considered not to have been made and the International Searching Authority will so declare.

4. ☒ Claim(s) Nos. -***-Please See Supplemental Page-***- have been found to be unsearchable under Article 17(2)(b) because of defects under Article 17(2)(a) and therefore have not been included with any invention.

Name and mailing address of the ISA/US: Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450	Authorized officer: Shane Thomas
Facsimile No. 571-273-8300	PCT Helpdesk: 571-272-4300

Form PCT/ISA/206 (July 2019)

INVITATION TO PAY ADDITIONAL FEES
AND, WHERE APPLICABLE, PROTEST FEE

International application No.
PCT/US19/65857

Item 2 (continued). For International Applications filed on or after 01 January 2014, Applicant is reminded that the search fee per additional invention indicated in item 2 is the undiscounted fee per additional invention. An Applicant may pay the search fee per additional invention fee reduced by 50% (small entity assertion) or 75% (micro entity certification), as appropriate. See 37 CFR 1.27 and 1.29.

-***-Continued from 1. (i)-***-

Group I, Claims 1-3, 100-114, 116-120, 122-133, 139-145, 148-152, 155-159, 162-166, 170-174, and 178-183
Group II, Claims 134-138
Group III, Claims 184-200 and 204-208

-***-Continued from 1. (ii)-***-

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, Claims 1-3, 100-114, 116-120, 122-133, 139-145, 148-152, 155-159, 162-166, 170-174, and 178-183 are directed toward methods comprising exposing a sample to nucleic acid probes, amplifying the probes, determining a distribution of the probes within the sample, creating codewords based on said distribution, and error correcting invalid codewords.

Group II, Claims 134-138 are directed toward a method comprising exposing a sample to nucleic acid probes, binding amplifying nucleic acids within a fixed distance to the probes, determining a distribution of the probes within the sample, and creating codewords based on said distribution.

Group III, Claims 184-200 and 204-208 are directed toward methods comprising exposing a sample to a binding or targeting entity conjugated to nucleic acid probes, binding nucleic acids to the probes, determining a distribution of the probes within the sample, and creating codewords based on said distribution.

The inventions listed as Groups I- and III do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical features of Group I include error correction, not present in either of Groups II or III; the special technical features of Group II include primary amplifier nucleic acids which bind to probes within a fixed distance, not present in either of Groups I or III; the special technical features of Group III include a targeting or binding entity, not present in either of Groups I or II.

Groups I-III share the technical features including: a method, comprising: exposing a sample to nucleic acid probes; exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes; determining a distribution of the nucleic acid probes within the sample using fluorescence; and creating codewords based on the fluorescence distribution within the sample. Groups I and III further share the technical features including: wherein the binding of primary amplifier nucleic acids and secondary amplifier nucleic acids to a target is saturatable; exposing primary amplifier nucleic acids to secondary amplifier nucleic acids able to bind to the primary amplifier nucleic acids; wherein the primary amplifier nucleic acids are formed from only 3 of the 4 naturally-occurring nucleotides; and wherein the secondary amplifier nucleic acids are formed from only 3 of the 4 naturally-occurring nucleotides.

However, these shared technical features are previously disclosed by WO 2018/026873 A1 to California Institute of Technology (hereinafter 'California') in view of US 6,150,516 A to Brenner et al. (hereinafter 'Brenner').

California discloses a method, comprising: exposing a sample to nucleic acid probes (a method, comprising: exposing a sample to nucleic acid probes; paragraph [0008]); exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes (exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes; paragraphs [00115], [00125], [00227]); determining a distribution of the nucleic acid probes within the sample using fluorescence (visually identifying probes bound to targets within a sample (determining a distribution of the nucleic acid probes within the sample using fluorescence); paragraph [0008]); and creating codewords based on the fluorescence distribution within the sample (and barcoding the sample based on serial probe hybrization and visualization rounds (creating codewords based on the fluorescence distribution within the sample); paragraph [0008]); wherein the binding of primary amplifier nucleic acids and secondary amplifier nucleic acids to a target is saturatable (wherein there are N molecular targets in the sample (wherein the binding of primary amplifier nucleic acids and secondary amplifier nucleic acids to a target is saturatable); paragraphs [0008], [00115], [00227]; wherein the limited number of targets naturally creates a saturatable level of target binding sites for probes and amplifier nucleic acids); and exposing primary amplifier nucleic acids to secondary amplifier nucleic acids able to bind to the primary amplifier nucleic acids (exposing secondary probes (primary amplifier nucleic acids) to tertiary probes (secondary amplifier nucleic acids) able to bind to the secondary probes (primary amplifier nucleic acids); paragraphs [00115], [00125], [00227]).

California does not disclose: wherein the primary amplifier nucleic acids are formed from only 3 of the 4 naturally-occurring nucleotides; and wherein the secondary amplifier nucleic acids are formed from only 3 of the 4 naturally-occurring nucleotides.

-***-Continued on Next Supplemental Page-***-

INVITATION TO PAY ADDITIONAL FEES
AND, WHERE APPLICABLE, PROTEST FEE

International application No.
PCT/US19/65857

-***-Continued from Previous Supplemental Page-***-:

Brenner discloses wherein minimally cross-hybridizing sets of oligonucleotide tags and complements thereof may be synthesized (wherein minimally cross-hybridizing sets of oligonucleotide tags and complements thereof may be synthesized; column 3, lines 36-40), depending on the size of the set desired and the degree to which cross-hybridization is sought to be minimized (depending on the size of the set desired and the degree to which cross-hybridization is sought to be minimized; column, 3, lines 36-40), using 3 of the 4 different types of nucleotides (using 3 of the 4 different types of nucleotides; column 3, lines 47-47, lines 56-59).

It would have been obvious to a person of ordinary skill in the art at the time the invention was made to have modified the disclosure of California to have used only 3 of the 4 different types of naturally occurring nucleotides for the formation of minimally cross-hybridizing sets of probes, based on the disclosure of Brenner, in order to better enable multiplexing of the detection of targets in a sample using a plurality of non-cross-hybridizing amplification probes, based on the disclosure of California, to more efficiently and effectively detect and barcode target sequences within the sample.

Since none of the special technical features of the Groups I-III inventions is found in more than one of the inventions, and since all of the shared technical features are previously disclosed by a combination of the California and Brenner references, unity of invention is lacking.

-***-Continued from 4.-***-

Claims 4-99, 115, 121, 146, 147, 153, 154, 160, 161, 167-169, 175-177, 201-203, 209-211

Electronic Acknowledgement Receipt

EFS ID:	39194787
Application Number:	15329683
International Application Number:	
Confirmation Number:	7428
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Customer Number:	86110
Filer:	Tani Chen/Catherine Borgnis
Filer Authorized By:	Tani Chen
Attorney Docket Number:	H0498.70509US02
Receipt Date:	17-APR-2020
Filing Date:	27-JAN-2017
Time Stamp:	16:41:03
Application Type:	U.S. National Stage under 35 USC 371

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	H049870509US02-TRN-TC.pdf	26190 519caf9f42688a132bcf858c72ec3d3c422a78bb	no	1
Warnings:					

Information:					
2	Transmittal Letter	H049870509US02-IDSBODY-TC.pdf	33761 7d846a42fea0de1e36f40a996cb58e7747f01f98	no	3
Warnings:					
Information:					
3	Information Disclosure Statement (IDS) Form (SB08)	H049870509US02-IDS1449-TC.pdf	23282 a4ffa068a8dbe8ceb9642ec0cc5e699aaf0d9889	no	1
Warnings:					
Information:					
This is not an USPTO supplied IDS fillable form					
4	Foreign Reference	WO2018026873A1.pdf	15685860 4c03ac708a62fe1206370e4c2d1e994a6f17904d	no	190
Warnings:					
Information:					
5	Other Reference-Patent/App/Search documents	H049870510CN00_CN_OA_mailed_03-05-2020.pdf	172487 1927c901bf9411cbcf4b78962b6859ba8d2860b5	no	7
Warnings:					
Information:					
6	Other Reference-Patent/App/Search documents	H049870637WO00_IPRP_mailed_12-5-19.pdf	569208 99e87d83d7b1f7a91344635af278f7ee7401026c	no	10
Warnings:					
Information:					
7	Other Reference-Patent/App/Search documents	H049870680WO00_Inv_to_Pay_mailed_03-03-2020.pdf	458277 5774489719f56fb6bfdd490c6461a96c17414509	no	3
Warnings:					
Information:					
8	Non Patent Literature	Gunderson_et_al_Methods_2004_v14_p870.pdf	291522 41b816d35bb3c04591e61aa25aaa2a30705c7445	no	8
Warnings:					
Information:					

9	Non Patent Literature	Epstein_et_al_Cytometry_1995_v21_p378_QC.pdf	323697 e77c0583e19ff60ff4dd1a471c6e81dfac1fc98f	no	4
Warnings:					
Information:					
			Total Files Size (in bytes):	17584284	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/329,683-Conf. #7428
Filing Date	January 27, 2017
First Named Inventor	Xiaowei Zhuang
Art Unit	1634
Examiner Name	F. W. Lu
Attorney Docket Number	H0498.70509US02

ENCLOSURES (*Check all that apply*)

- [ ] Fee Transmittal Form
  - [ ] Fee Attached
- [ ] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [x] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53

- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD

- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) (please Identify below):

Form PTO-1449
Copies of Cited References

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Tani Chen/		
Printed name	Tani Chen, Sc.D.		
Date	April 17, 2020	Reg. No.	52,728

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: April 17, 2020 Electronic Signature for Catherine E. Borgnis: _/Catherine E. Borgnis/_

8134926.1

DOCKET NO.: H0498.70509US02

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	Xiaowei Zhuang
Application No.:	15/329,683
Confirmation No.:	7428
Filed:	January 27, 2017
For:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
Examiner:	F. W. Lu
Art Unit:	1634

Certificate of Electronic Filing under 37 C.F.R. § 1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: April 17, 2020 Electronic Signature for Catherine E. Borgnis: _/Catherine E. Borgnis/_

MAIL STOP RCE
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

STATEMENT FILED PURSUANT TO THE DUTY OF DISCLOSURE UNDER 37 C.F.R. §§ 1.56, 1.97 AND 1.98

Sir:

Pursuant to the duty of disclosure under 37 C.F.R. §§ 1.56, 1.97 and 1.98, the undersigned requests consideration of this Information Disclosure Statement.

PART I: Compliance with 37 C.F.R. § 1.97

This Information Disclosure Statement has been filed before the mailing of a first Office Action after the filing of a Request for Continued Examination under 37 C.F.R. § 1.114.

No fee or certification is required.

PART II: Information Cited

The undersigned hereby makes of record in the above-identified application the information listed on the attached form PTO-1449 (modified PTO/SB/08). The order of presentation of the references should not be construed as an indication of the importance of the references.

The undersigned hereby makes the following additional information of record in the above-identified application.

The undersigned would like to bring to the Examiner's attention the following co-pending applications that may contain subject matter related to this application:

Serial No.	Filing Date	Inventor(s)	Docket No.
*16/253,919	1/22/2019	Zhuang et al.	H0498.70510US02
*16/347,874	5/7/2019	Zhuang et al.	H0498.70630US01
*16/616,833	11/25/2019	Zhuang et al.	H0498.70637US01
*16/348,071	5/7/2019	Zhuang et al.	H0498.70645US01

*A copy of this reference is not provided as the Office has waived the requirement under 37 C.F.R. § 1.98(a)(2)(iii) for submitting a copy of a cited U.S. patent application if it is scanned to the Image File Wrapper system and is available on Private PAIR .

The undersigned would like to bring to the Examiner's attention the enclosed search report or other communication from a corresponding or related International or Foreign National Application:

Serial No.	Date of Mailing	Type(s) of Communication	Docket No.
CN 201580051029.5	**03-05-2020**	**Chinese Office Action**	**H0498.70510CN00**
PCT/US18/34651	**12-05-2019**	**International Preliminary Report on Patentability**	**H0498.70637WO00**
PCT/US2019/065857	**03-03-2020**	**Invitation to Pay Additional Fees**	**H0498.70680WO00**

PART III: Remarks

Documents cited anywhere in the Information Disclosure Statement are enclosed unless otherwise indicated. It is respectfully requested that:

1. The Examiner consider completely the cited information, along with any other information, in reaching a determination concerning the patentability of the present claims;

2. The enclosed form PTO-1449 (modified PTO/SB/08) be signed by the Examiner to evidence that the cited information has been fully considered by the United States Patent and Trademark Office during the examination of this application;

3. The citations for the information be printed on any patent which issues from this application.

By submitting this Information Disclosure Statement, the undersigned makes no representation that a search has been performed, of the extent of any search performed, or that more relevant information does not exist.

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, material to patentability as defined in 37 C.F.R. § 1.56(b).

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, in fact, prior art as defined by 35 U.S.C. § 102.

Notwithstanding any statements by the undersigned, the Examiner is urged to form his or her own conclusion regarding the relevance of the cited information.

An early and favorable action is hereby requested.

The Director is hereby authorized to charge any deficiency or credit any overpayment in the fees occasioned by the filing of this Information Disclosure Statement to our Deposit Account No. 23/2825 under Docket No. H0498.70509US02 from which the undersigned is authorized to draw.

Respectfully submitted,

Electronic signature: /Tani Chen/
Tani Chen, Sc.D.
Registration No.: 52,728
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

Docket No.: H0498.70509US02
Date: April 17, 2020

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/329,683	01/27/2017	Xiaowei Zhuang	H0498.70509US02	7428

86110 7590 10/22/2020

Wolf, Greenfield & Sacks, P.C. (Harvard)
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER	
LU, FRANK WEI MIN	

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
10/22/2020	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

H0498_eOfficeAction@WolfGreenfield.com
Patents_eOfficeAction@WolfGreenfield.com

Office Action Summary	Application No. 15/329,683	Applicant(s) Zhuang et al.	
	Examiner FRANK W LU	Art Unit 1634	AIA (FITF) Status Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☑ Responsive to communication(s) filed on 4/17/2020.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☐ This action is **FINAL.** 2b) ☑ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5) ☑ Claim(s) 1,34,71,75,77-78 and 177-182 is/are pending in the application.

5a) Of the above claim(s) _____ is/are withdrawn from consideration.

6) ☐ Claim(s) _____ is/are allowed.

7) ☑ Claim(s) 1,34,71,75,77-78 and 177-182 is/are rejected.

8) ☐ Claim(s) _____ is/are objected to.

9) ☐ Claim(s) _____ are subject to restriction and/or election requirement

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to **PPHfeedback@uspto.gov.**

Application Papers

10)☐ The specification is objected to by the Examiner.

11)☑ The drawing(s) filed on 7/29/2019 is/are: a)☑ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☑ Notice of References Cited (PTO-892)

2) ☑ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date 10/31/2019, 12/11/2019, and 4/17/2020.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date _____.

4) ☐ Other: _____.

DETAILED ACTION

CONTINUED EXAMINATION UNDER 37 CFR 1.114 AFTER FINAL REJECTION

1. A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission of RCE and the amendment filed on April 17, 2020 have been entered. The claims pending in this application are claims 1, 34, 71, 75, 77, 78, and 177-182. Rejection and/or objection not reiterated from the previous office action are hereby withdrawn in view of applicant's amendment filed on April 17, 2020. Claims 1, 34, 71, 75, 77, 78, and 177-182 will be examined.

Claim Objections

2. Claim 1 is objected to because of the following informalities: (1) "a plurality of nucleic acids targets" in (ii) of step (a) should be "a plurality of nucleic acid targets"; (2) "each nucleic acid target" in (ii) of step (a) should be "each of the nucleic acid targets"; and (3) "read sequences and codewords" should be "the read sequences and the codewords".

3. Claim 34 is objected to because of the following informality: "after exposing the sample to one or more readout nucleic acid probes and imaging the sample to determine the fluorescence signal of the fluorescent signaling entities" should be "after said exposing the sample to the one or more readout nucleic acid probes and imaging the sample to determine the fluorescence signal of the fluorescent signaling entities".

4. Claim 180 is objected to because of the following informality: "unbound readout nucleic acid probes prior to imaging the sample" should be "the readout nucleic acid probes unbound prior to said imaging the sample".

5. Claim 181 or 182 is objected to because of the following informality: "the acts of exposing the sample to one or more readout nucleic acid probes, imaging the sample, and inactivating at least some of the fluorescent signaling entities on the read out probes" should be "said exposing the sample to one or more readout nucleic acid probes, said imaging the sample, and said inactivating at least some of the fluorescent signaling entities on the read out probes".

Appropriate correction is required.

Claim Rejections - 35 USC § 112

6. The following is a quotation of the first paragraph of 35 U.S.C. 112(a):

> (a) IN GENERAL.—The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor or joint inventor of carrying out the invention.

The following is a quotation of the first paragraph of pre-AIA 35 U.S.C. 112:

> The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor of carrying out his invention.

7. Enablement

Claims 1, 34, 71, 75, 77, 78, and 177-182 are rejected under 35 U.S.C. 112(a) or 35 U.S.C. 112 (pre-AIA), first paragraph, as failing to comply with the enablement requirement. The claim(s) contains subject matter which was not described in the specification in such a way

as to enable one skilled in the art to which it pertains, or with which it is most nearly connected, to make and/or use the invention.

In *In re Wands*, 858 F.2d 731,737, 8 USPQ2d 1400, 1404 (Fed. Cir. 1988) the court considered the issue of enablement in molecular biology. The Court summarized eight factors to be considered in a determination of "undue experimentation". These factors include: (a) the quantity of experimentation necessary; (b) the amount of direction or guidance presented; (c) the presence or absence of working examples; (d) the nature of the invention; (e) the state of the prior art; (f) the relative skill of those in the art; (g) the predictability of the art; and (h) the breadth of the claims. The Court also stated that although the level of skill in molecular biology is high, results of experiments in molecular biology are unpredictable.

To begin, there is no direction or guidance to show that the methods recited in claims 1, 34, 71, 75, 77, 78, and 177-182 can be performed. While the relative skill in the art is very high (the Ph.D. degree with laboratory experience), there is no predictability whether the methods recited in claims 1, 34, 71, 75, 77, 78, and 177-182 can be performed.

Since step (c) of claim 1 does not indicate that each of a first round of one or more readout nucleic acid probes only specifically hybridizes to one unique reading sequence of a plurality of reading sequences and each of the first round of one or more readout nucleic acid probes contains a different fluorescent signaling entity, if each of the first round of one or more readout nucleic acid probes can nonspecifically hybridize to two or more unique reading sequences of the plurality of reading sequences and each of the first round of one or more readout nucleic acid probes contains an identical fluorescent signaling entity, the reading sequences of the plurality of reading sequences identified by each of the first round of one or more readout nucleic acid probes are non-specific and a fluorescence signal produced from the

fluorescent signaling entity are non-specific such that the read sequences and the codewords at the locations in the sample cannot be determined based on the fluorescence signal, substeps (iv) and (v) of step (c) of claim 1 cannot performed and an abundance and/or spatial distribution of the nucleic acid targets recited in claim 179 cannot be determined. Furthermore, it is unclear that, in which situation, the read sequences and the codewords at the locations can be determined based on the fluorescence signal and in which situation, a codeword matches with a valid codeword and in which situation, a codeword does not match with a valid codeword. In addition, since claim 1 does not indicate how a nucleic acid abundance and/or spatial distribution within the sample is correlate with valid codewords, it is unclear that, by which way, a nucleic acid abundance and/or spatial distribution within the sample can be determined.

In view of above discussions, the skilled artisan will have no way to predict the experimental results. Accordingly, it is concluded that undue experimentation is required to make the invention as it is claimed. These undue experimentation at least includes to test whether the methods recited in claims 1, 34, 71, 75, 77, 78, and 177-182 can be performed.

8. The following is a quotation of 35 U.S.C. 112(b):
(b) CONCLUSION.—The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the inventor or a joint inventor regards as the invention.

The following is a quotation of 35 U.S.C. 112 (pre-AIA), second paragraph:
The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

9. Claims 1, 34, 71, 75, 77, 78, and 177-182 are rejected under 35 U.S.C. 112(b) or 35 U.S.C. 112 (pre-AIA), second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which the inventor or a joint inventor (or for applications subject to pre-AIA 35 U.S.C. 112, the applicant), regards as the invention.

10. Claim 1 is rejected as vague and indefinite in view of substep (ii) of step (a) because it is unclear what is the relationship between the valid codewords and a unique valid codeword. Please clarify.

11. Claim 1 is rejected as vague and indefinite in view of substep (i) of step (b) because it is unclear what is the relationship between a plurality of nucleic acid read sequences and a unique read sequence. Please clarify.

12. Claim 181 is rejected as vague and indefinite because it is unclear that the acts of exposing the sample to one or more readout nucleic acid probes, imaging the sample, and inactivating at least some of the fluorescent signaling entities on the read out probes are performed for what purpose related to each of the values, or subset thereof, at each of the positions within the codeword space. Please clarify.

13. Claim 182 is rejected as vague and indefinite because it is unclear that the acts of exposing the sample to one or more readout nucleic acid probes, imaging the sample, inactivating at least some of the fluorescent signaling entities on the readout probes, and assigning codeword position values are performed for what purpose related to each of the values, or subset thereof, at each of the positions within the codeword space. Please clarify.

Claim Rejections - 35 USC § 101

14. 35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

15. Claims 1, 34, 71, 75, 77, 78, and 177-182 are rejected under 35 U.S.C. 101 because the claimed invention is directed to a judicial exception (i.e., a law of nature, a natural phenomenon, or an abstract idea) without significantly more.

The processes of claims 1, 34, 71, 75, 77, 78, and 177-182 are directed to a process of identifying nucleic acid targets in the sample based on the codewords matched to the valid codewords. Claim 1 includes a) associating a plurality of nucleic acid targets with valid codewords, comprising: (i) defining a codeword space comprising a plurality of valid codewords, wherein the valid codewords in the codeword space are defined by a series of positions and values for each at the positions, and the valid codewords form an error-checking and/or error-correcting code space; and (ii) associating the valid codewords with a plurality of nucleic acids targets such that each nucleic acid target is associated with a unique valid codeword; (b) exposing a sample suspected of containing at least some of the plurality of nucleic acid targets to a plurality of encoding nucleic acid probes that contain a representation of the valid codewords, comprising: (i) providing a plurality of nucleic acid read sequences and associating a unique read sequence with each of the values, or a subset thereof, at each of the positions of the valid codewords; (ii) creating a plurality of encoding nucleic acid probes, wherein each of the encoding nucleic acid probes comprises a first portion comprising a target sequence able to bind to one of the nucleic acid targets via nucleic acid hybridization, and a second portion comprising one or more read sequences taken from the plurality of read sequences; and (iii) exposing the sample to the plurality of encoding nucleic acid probes to cause the encoding nucleic acid probes to bind to at least some of the nucleic acid targets in the sample via nucleic acid hybridization; and (c) identifying nucleic acid targets in the sample by determining valid codewords that correspond to the plurality of nucleic acid probes, comprising:

(i) exposing the sample to a first round of one or more readout nucleic acid probes, wherein at least some of the readout nucleic acid probes comprise a first portion and a second portion, the first portion comprising a sequence substantially complementary to one of the read sequences, and the second portion comprising a fluorescent signaling entity; (ii) imaging the sample to determine a fluorescence signal of the fluorescent signaling entity in at least some locations in the sample; (iii) determining read sequences and codewords at the locations based on the fluorescence signal; (iv) matching the codewords to the valid codewords in the codeword space wherein, if no match is found, either discarding the codeword or applying error correction to the codeword to match one of the valid codewords; and (v) identifying nucleic acid targets in the sample based on the codewords matched to the valid codewords, wherein step (a), substep (i) of step (b), and substeps (iv) and (v) of step (c) in independent claim 1 fall within the mental process grouping of abstract ideas. **Note that steps in a claim that recite the manipulation of information through a series of mental steps would be considered a single abstract idea for purposes of analysis** (see MPEP 2106.04).

This judicial exception, the abstract idea, is not integrated into a practical application because the additional elements, substeps (ii) and (iii) of step (b) and substeps (i) to (iii) of step (c) of claim 1 do not integrate the mental steps comprising substep (i) of step (b), and substeps (iv) and (v) of step (c) of claim 1 into a practical application.

The claims do not include additional elements that are sufficient to amount to significantly more than the judicial exception because, since, in independent claim 1, substep (ii) of step (b) only requires creating a plurality of encoding nucleic acid probes, wherein each of the encoding nucleic acid probes comprises a first portion comprising a target sequence able to bind

to one of the nucleic acid targets via nucleic acid hybridization, and a second portion comprising one or more read sequences taken from the plurality of read sequences, substep (iii) of step (b) only requires exposing the sample to the plurality of encoding nucleic acid probes to cause the encoding nucleic acid probes to bind to at least some of the nucleic acid targets in the sample via nucleic acid hybridization, and steps (i) to (iii) of step (c) only requires exposing the sample to a first round of one or more readout nucleic acid probes, wherein at least some of the readout nucleic acid probes comprise a first portion and a second portion, the first portion comprising a sequence substantially complementary to one of the read sequences, and the second portion comprising a fluorescent signaling entity, imaging the sample to determine a fluorescence signal of the fluorescent signaling entity in at least some locations in the sample, and determining read sequences and codewords at the locations based on the fluorescence signal, and the detection techniques used in substeps (ii) and (iii) of step (b) and substeps (i) to (iii) of step (c) of claim 1 are common, well-understood, routine and conventional methods used in the laboratory and are well known in the art (for example, see columns 1-5, 7-9, and 19, Figures 1-3, and claims 1-15). The conventional steps do not comprise an inventive concept that transforms the claimed abstract ideas into a patent-eligible application of the abstract ideas. The claimed subject matters of claims 1, 34, 71, 75, 77, 78, and 177-182 are processes directed to abstract ideas. When recited at this high level of generality, there is no meaningful limitation, such as a particular or unconventional machine or a transformation of a particular article, in this step that distinguishes it from well-understood, routine, and conventional data gathering activity engaged in by scientists prior to applicant's invention, and at the time the application was filed, *e.g.*, the routine and conventional techniques for performing substeps (ii) and (iii) of step (b) and substeps (i) to (iii) of step (c) of claim 1. Further, it is well established that the mere physical or tangible nature

of additional elements such as substeps (ii) and (iii) of step (b) and substeps (i) to (iii) of step (c) of claim 1 do not automatically confer eligibility on a claim directed to an abstract idea (*see*, *e.g.*, Alice Corp. v. CLS Bank Int'l, 134 S. Ct. 2347, 2358-59 (2014)). Therefore, the claimed subject matters of claims 1, 34, 71, 75, 77, 78, and 177-182 are without limitations that indicate that the claims are no different from the process claims directed to abstract ideas and are patent-ineligible.

Response to Arguments

16. Applicant's arguments with respect to claim 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78 have been considered but are moot because the new ground of rejection does not rely on any reference applied in the prior rejection of record for any teaching or matter specifically challenged in the argument.

Conclusion

17. No claim is allowed.

18. Papers related to this application may be submitted to Group 1600 by facsimile transmission. Papers should be faxed to Group 1600 via the PTO Fax Center. The faxing of such papers must conform with the notices published in the Official Gazette, 1096 OG 30 (November 15, 1988), 1156 OG 61 (November 16, 1993), and 1157 OG 94 Fax Center number is (571)273-8300.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Frank Lu, Ph.D., whose telephone number is (571)272-0746. The examiner can normally be reached on Monday-Friday from 9 A.M. to 5 P.M.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Dave Nguyen, can be reached on (571)272-0731.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/FRANK W LU/
Primary Examiner, Art Unit 1634
October 1, 2020

Notice of References Cited	Application/Control No. 15/329,683	Applicant(s)/Patent Under Reexamination Zhuang et al.	
	Examiner FRANK W LU	Art Unit 1634	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	CPC Classification	US Classification
*	A	US-10501791-B2	12-2019	Church; George M.	C12Q1/6874	1/1
	B					
	C					
	D					
	E					
	F					
	G					
	H					
	I					
	J					
	K					
	L					
	M					

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	CPC Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15/329,683	Zhuang et al.
	Examiner	Art Unit
	FRANK W LU	1634

CPC - Searched*		
Symbol	**Date**	**Examiner**
Update	10/01/2020	/FL/

CPC Combination Sets - Searched*		
Symbol	**Date**	**Examiner**
Update	10/01/2020	/FL/

US Classification - Searched*			
Class	**Subclass**	**Date**	**Examiner**
Update		10/01/2020	/FL/

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

Search Notes		
Search Notes	**Date**	**Examiner**
Update	10/01/2020	/FL/

Interference Search			
US Class/CPC Symbol	**US Subclass/CPC Group**	**Date**	**Examiner**

/FRANK W LU/ Primary Examiner, Art Unit 1634	October 1, 2020

FORM PTO-1449/A and B (modified PTO/SB/08)				APPLICATION NO.: 15/329,683	ATTY. DOCKET NO.: H0498.70509US02
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				FILING DATE: January 27, 2017	CONFIRMATION NO.: 7428
				FIRST NAMED INVENTOR: Xiaowei Zhuang	
Sheet	1	of	1	GROUP ART UNIT: 1634	EXAMINER: F. W. Lu

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document		Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		Number	Kind Code		

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document			Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)
		Office/ Country	Number	Kind Code			

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		European Office Action mailed October 16, 2019 for Application No. 15827358.1 (H0498.70509EP00).	

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

EXAMINER:	DATE CONSIDERED:
/FRANK W LU/	10/01/2020

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

7681167.1

FORM PTO-1449/A and B (modified PTO/SB/08)				APPLICATION NO.: 15/329,683	ATTY. DOCKET NO.: H0498.70509US02
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				FILING DATE: January 27, 2017	CONFIRMATION NO.: 7428
				FIRST NAMED INVENTOR: Xiaowei Zhuang	
Sheet	1	of	1	GROUP ART UNIT: 1634	EXAMINER: F. W. Lu

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document Number	Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		6,150,516	A1	Brenner et al.	11-21-2000
		7,226,734	B2	Chee et al.	06-05-2007
		2006-0275782	A1	Gunderson et al.	12-07-2006
		2015-0267251	A1	Cai et al.	09-24-2015
		2019-0233812	A1	Zhuang et al.	08-01-2019
		2019-0264270	A1	Zhuang et al.	08-29-2019
		2019-0276881	A1	Zhuang et al.	09-12-2019
		2020-0095630	A1	Zhuang et al.	03-26-2020

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document Office/ Country	Number	Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)
		WO	2018/026873	A1	California Institute of Technology	02-08-2018	

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		Chinese Office Action mailed March 5, 2020 for Application No. CN 201580051029.5 (H0498.70510CN00).	
		International Preliminary Report on Patentability mailed December 5, 2019 for Application No. PCT/US18/34651 (H0498.70637WO00).	
		Invitation to Pay Additional Fees for Application No. PCT/US2019/065857 mailed March 3, 2020 (H0498.70680WO00).	
		EPSTEIN et al., Reutilization of Previously Hybridized Slides for Fluorescence In Situ Hybridization. Cytometry. 1995;21:378-381.	
		GUNDERSON et al., Decoding Randomly Ordered DNA Arrays. Methods. 2004;14:870–877.	

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

EXAMINER:	DATE CONSIDERED:
/FRANK W LU/	10/01/2020

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

8134927.1

FORM PTO-1449/A and B (modified PTO/SB/08)				APPLICATION NO.: 15/329,683	ATTY. DOCKET NO.: H0498.70509US02
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				FILING DATE: January 27, 2017	CONFIRMATION NO.: 7428
				FIRST NAMED INVENTOR: Xiaowei Zhuang	
Sheet	1	of	1	GROUP ART UNIT: 1634	EXAMINER: F. W. Lu

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document		Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		Number	Kind Code		

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document			Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)
		Office/ Country	Number	Kind Code			

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		Chinese Office Action dated November 14, 2019 for Application No. 201580052678.7 (H0498.70509CN00).	Y - Full

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

EXAMINER:	DATE CONSIDERED:
/FRANK W LU/	10/01/2020

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

7786753.1

Docket No.: H0498.70509US02
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	Xiaowei Zhuang
Application No.:	15/329,683
Confirmation No.:	7428
Filed:	January 27, 2017
For:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
Examiner:	F. W. Lu
Art Unit:	1634

Certificate of Electronic Filing under 37 C.F.R. § 1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: January 22, 2021 Electronic Signature for Rachel Moulton: /Rachel Moulton/

RESPONSE

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

INTRODUCTORY COMMENTS

In response to the Office Action dated October 22, 2020, reconsideration is respectfully requested in view of the following:

The Claims are reflected in the listing of claims which begins on page 2 of this paper.

Remarks begin on page 7 of this paper.

Electronic Acknowledgement Receipt

EFS ID:	41713953
Application Number:	15329683
International Application Number:	
Confirmation Number:	7428
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Customer Number:	86110
Filer:	Tani Chen/Rachel Moulton
Filer Authorized By:	Tani Chen
Attorney Docket Number:	H0498.70509US02
Receipt Date:	22-JAN-2021
Filing Date:	27-JAN-2017
Time Stamp:	09:32:19
Application Type:	U.S. National Stage under 35 USC 371

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Transmittal Letter	H049870509US02-TRN-TC.pdf	26097 16359d9f5d53af7a50e9bcf8f57c86c8f561062b	no	1

Warnings:

Information:					
2		H049870509US02-AMN-TC.pdf	62133 3b116ad60592cf9c39a528f6c0c110472e545a71	yes	13

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Applicant Arguments/Remarks Made in an Amendment	7	13
	Claims	2	6
	Amendment/Req. Reconsideration-After Non-Final Reject	1	1

Warnings:	
Information:	
Total Files Size (in bytes):	88230

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Doc Code: TRAN.LET
Document Description: Transmittal Letter

USED IN LIEU OF PTO/SB/21 (07-09)

TRANSMITTAL FORM *(to be used for all correspondence after initial filing)*			
		Application Number	15/329,683-Conf. #7428
		Filing Date	January 27, 2017
		First Named Inventor	Xiaowei Zhuang
		Art Unit	1634
		Examiner Name	F. W. Lu
Total Number of Pages in This Submission		Attorney Docket Number	H0498.70509US02

ENCLOSURES (*Check all that apply*)

☐ Fee Transmittal Form	☐ Drawing(s)	☐ After Allowance Communication to TC
☐ Fee Attached	☐ Licensing-related Papers	☐ Appeal Communication to Board of Appeals and Interferences
☒ Amendment/Reply	☐ Petition	☐ Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
☐ After Final	☐ Petition to Convert to a Provisional Application	☐ Proprietary Information
☐ Affidavits/declaration(s)	☐ Power of Attorney, Revocation Change of Correspondence Address	☐ Status Letter
☐ Extension of Time Request	☐ Terminal Disclaimer	☐ Other Enclosure(s) (please Identify below):
☐ Express Abandonment Request	☐ Request for Refund	
☐ Information Disclosure Statement	☐ CD, Number of CD(s) _____	
☐ Certified Copy of Priority Document(s)	☐ Landscape Table on CD	
☐ Reply to Missing Parts/ Incomplete Application	Remarks	
☐ Reply to Missing Parts under 37 CFR 1.52 or 1.53		

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Tani Chen/		
Printed name	Tani Chen, Sc.D.		
Date	January 22, 2021	Reg. No.	52,728

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: January 22, 2021 Electronic Signature for Rachel Moulton: /Rachel Moulton/

8647840.1

REMARKS

Claim 182 has been cancelled without prejudice or disclaimer. Claim 1, 34, and 177-181 have been amended to make the claim language even more clear and address the concerns raised by the Patent Office as discussed below. As it may be difficult to read claim 1, it is reproduced below in clean form, without strikethroughs and underlines:

1. A method for detecting nucleic acid targets, comprising:
(a) exposing a sample suspected of containing at least some of a plurality of nucleic acid targets to a plurality of encoding nucleic acid probes,
wherein at least some of the plurality of encoding nucleic acid probes comprise a target sequence that specifically hybridizes to a nucleic acid target in the plurality of nucleic acid targets and one or more read sequences,
wherein unique combinations of one or more read sequences correspond to valid codewords, each of the valid codewords associated to each of the plurality of nucleic acid targets, wherein the valid codeword consists of a series of positions and values for each position, and
each read sequence is associated with each of the values, or a subset thereof, at each of the positions of the valid codewords;
(b) exposing the sample to a first round of one or more readout nucleic acid probes, wherein at least some of the readout nucleic acid probes comprise a first portion and a second portion, the first portion comprising a sequence that specifically hybridizes to one of the read sequences, and the second portion comprising a fluorescent signaling entity;
(c) identifying nucleic acid targets in the sample, comprising:
detecting fluorescent signals emitted from the fluorescent signaling entities of bound readout nucleic acid probes in at least some locations within the sample to determine a codeword in at least some locations within the sample based on the read sequences present on the bound encoding nucleic acid probes;
identifying the codeword at each location by comparing the codeword to the valid codewords associated with the plurality of nucleic acid targets, and if no match is found, applying error correction to the codeword to match one of the valid codewords or discarding the codeword; and
assigning the identity of some of the plurality of nucleic acid targets in the sample based on the identity of the nucleic acid targets associated with the valid codewords to which the identified codewords were matched.

New claims 183-187 have been added. Support for the new claims can be found in the as-filed specification, e.g., page 25, lines 5-15; page 26 line 25 to page 27, line 11; Example 4, page 59.

No new matter has been added by these amendments and new claims. Claims 1, 34, 71, 75, 77, 78, 177-181, and 183-187 are now pending for examination.

Claim Objections

Claims 1, 34, 180, 181, and 182 have been objected to for the alleged antecedent basis issues. Office Action at pages 2-3.

These claims have been amended as shown above, generally adopting the Examiner's suggestions. Accordingly, withdrawal of these objections is respectfully requested.

Rejections under 35 U.S.C. § 112(a)

Claims 1, 34, 71, 75, 77, 78, and 177-182 have been rejected under 35 U.S.C. § 112(a) because according to the Patent Office, the specification, while being enabling for exposing a sample to a plurality of nucleic acid probes, does not reasonably provide enablement for performing the methods recited in claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78.

In particular, the Patent Office alleges on pages 4-5 of the Office Action that:

> step (c) of claim 1 does not indicate that each of a first round of one more readout acid probes only specifically hybridizes to one unique reading sequence of a plurality of reading sequences and each of the first round of one or more readout nucleic acid contains a different fluorescent signaling entity if each of the first round of one or more readout nucleic acid probes can nonspecifically hybridize to two or more unique reading sequences of the plurality or reading sequences and each of the first round of one or more readout nucleic acid probes contains an identical fluorescent signaling entity, the reading sequences of the plurality of reading sequences identified by each of the first round of one or more readout nucleic acid probes are non-specific and a fluorescence signal produced from the fluorescent signaling entity are nonspecific such that the read sequences and the codewords at the locations in the sample cannot be determined based on the fluorescence signal, substeps (iv) and (v) of step (c) of claim 1 cannot performed and an abundance and/or spatial distribution of the nucleic acid targets recited in claim 179 cannot be determined.

Solely to expedite prosecution without acquiescing the rejection, claim 1 has been amended to recite "a first round of one or more readout nucleic acid probes, wherein at least some of the

readout nucleic acid probes comprise a target sequence that specifically hybridizes to a nucleic acid target in the plurality of nucleic acid targets, and the second portion comprising a fluorescent signaling entity" (emphasis added).

The Patent Office also alleges on page 5 of the Office Action that:

> it is unclear that, in which situation, the read sequences and the codewords at the locations can be determined based on the fluorescence signal and in which situation, a codeword matches with a valid codeword and in which situation, a codeword does not match with a valid codeword.

Solely to expedite prosecution without acquiescing the rejection, claim 1 has been amended to recite:

> detecting fluorescent signals emitted from the fluorescent signaling entities of bound readout nucleic acid probes in at least some locations within the sample to determine a codeword in at least some locations within the sample based on the read sequences present on the bound readout nucleic acid probes; and
>
> identifying the codeword at each location matched to one of the valid codewords associated with the plurality of nucleic acid targets, by comparing the values at each position of the codeword to the values at each position of the valid codeword, and if no match is found, applying error correction to the codeword to match one of the valid codewords or discard the codeword; and
>
> assigning the identity of some of the plurality of nucleic acid targets in the sample based on the identity of the nucleic acid targets associated with the valid codewords to which the identified codewords were matched.

The Patent Office also alleges on page 5 of the Office Action that:

> since claim 1 does not indicate how a nucleic acid abundance and/or spatial distribution within the sample is correlate with valid codewords, it is unclear that, by which way, a nucleic acid abundance and/or spatial distribution within the sample can be determined.

Claim 1, as previously amended, does not include the alleged limitations; however, claim 179 does. Claim 179 has been amended to recite "determining an abundance and/or spatial

distribution of the nucleic acid targets based on the codewords matched to the valid codewords in each location within the sample."

Thus, for at least these reasons, withdraw of the rejection of claims 1, 34, 71, 75, 77, 78, and 177-182 under 35 U.S.C. § 112(a) is respectfully requested.

Rejections under 35 U.S.C. § 112(b)

Claims 1, 34, 71, 75, 77, 78, and 177-182 have been rejected under 35 U.S.C. § 112(b), for allegedly being indefinite. Specifically, the Patent Office alleges that it is unclear what the relationship is between the valid codewords and a unique valid codeword, the terms recited in claim 1. The Patent Office also alleges that it is unclear what the relationship is between a plurality of nucleic acid read sequences and a unique read sequence, the terms recited in claim 1. Office Action at page 6.

Solely to expedite prosecution without acquiescing the rejection, claim 1 has amended to remove the alleged limitations and further clarify that "unique combinations of one or more read sequences correspond to valid codewords, each of the valid codewords associated to each of the plurality of nucleic acid targets, wherein the valid codeword consists of a series of positions and values for each position, and each read sequence is associated with each of the values, or a subset thereof, at each of the positions of the valid codewords."

Regarding claims 181 and 182, the Patent Office alleges that it is unclear that the recited steps are performed for what purpose related to each of the values, or subset thereof, at each of the positions within the codeword space. Office Action at page 6.

Solely to expedite prosecution without acquiescing the rejection, claim 181 has been amended to explicitly recite that the recited steps are performed "to assign each of the values, or subset thereof, at each of the positions within the codewords." Claim 182 has been cancelled without prejudice or disclaimer.

Thus, for at least these reasons, withdraw of the rejection of claims 1, 34, 71, 75, 77, 78, and 177-182 under 35 U.S.C. § 112(b) is respectfully requested.

Rejections under 35 U.S.C. § 101

Claims 1, 34, 71, 75, 77, 78, and 177-182 have been rejected under 35 U.S.C. § 101 because the claimed invention is directed to a judicial exception (i.e., a law of nature, a natural phenomenon, or an abstract idea).

Specifically, the Patent Office alleges that step (a), substep (i) of step (b), and substeps (iv) and (v) of step (c) in claim 1 fall within the mental process grouping of abstract ideas. The Patent Office further notes that steps in a claim that recite the manipulation of information through a series of mental steps would be considered a single abstract idea for purposes of analysis, accordingly to MPEP § 2106.04. Office Action at page 8. The Patent Office also alleges that "this judicial exception, the abstract idea, is not integrated into a practical application because the additional elements, substeps (ii) and (iii) of step (b) and substeps (i) to (iii) of step (c) of claim 1 do not integrate the mental steps comprising substep (i) of step (b), and substeps (iv) and (v) of step (c) in claim 1 into a practical application." *Id.* The Patent Office further alleges that the techniques used in substeps (ii) and (iii) of step (b) and substeps (i) to (iii) of step (c) of claim 1 are common, well-understood, routine and conventional methods used in the laboratory and are well known in the art, referring to columns 1-5, 7-9, and 19, Figures 1-3, and claims 1-15 of US Pat. No. 10,501,791 (as cited in the PTO-892 form attached to the Office Action) and alleges that the conventional steps do not comprise an inventive concept that transforms the claimed abstract ideas into a patent-eligible application of the abstract ideas. *Id.* at pages 9-10.

Solely to expedite prosecution and without acquiescing the rejection, claim 1 has been amended to remove the alleged substep (i) of step (b), and substeps (iv) and (v) of step (c) that the Patent Office alleges as mental steps. Claim 1 has also been amended to further clarify the features of unique combination of read sequences, encoding probes, and codewords associated therewith that allow error-detection in identifying nucleic acid targets. Amended claim 1 recites, among other things:

> exposing a sample suspected of containing at least some of the plurality of nucleic acid targets to a plurality of encoding nucleic acid probes, wherein at least some of the plurality of encoding nucleic acid probes comprise a target sequence that specifically hybridizes to a nucleic acid target in the plurality of nucleic acid targets and one or more read sequences, wherein unique combinations of one or more read sequences correspond to valid

codewords, each of the valid codewords associated to each of the plurality of nucleic acid targets, wherein the valid codeword consists of a series of positions and values for each position, each read sequence in the combinations is associated with each of the values, or a subset thereof, at each of the positions of the valid codewords;

...

identifying the codeword at each location matched to one of the valid codewords associated with the plurality of nucleic acid targets, by comparing the values at each position of the codeword to the values at each position of the valid codeword, and if no match is found, applying error correction to the codeword to match one of the valid codewords or discarding the codeword; and

assigning the identity of some of the plurality of nucleic acid targets in the sample based on the identity of the nucleic acid targets associated with the valid codewords to which the identified codewords were matched.

Arguendo, if the Patent Office were to allege these steps being mental steps, the amended claim would be still patent eligible for at least the following reasons.

Mere recitation of a judicial exception does not mean that the claim is "directed to" that judicial exception under Step 2A Prong Two; instead, a claim that recites a judicial exception is not directed to that judicial exception if the claim as a whole "integrates the recited judicial exception into a practical application of that exception." See USPTO October 2019 Update: Subject Matter Eligibility, page 10 and MPEP 2106.04(II). An important consideration to evaluate when determining whether the claim as a whole integrates a judicial exception into a practical application is whether the claimed invention improves upon conventional technology. October 2019 Update, page 12. The improvement can be provided by one or more additional elements, or by the additional element(s) in combination with the recited judicial exception. *Id.*, page 13.

Here, the totality of these steps including the recitation of particular unique combinations of read sequences and determination of the codewords based on the read sequences integrate the exception into the practical application, and amount to more than routine and conventional methods involving probe hybridization. Further, the combination of steps, which is not routine and conventional, ensures that the nucleic acid targets will be identified error-free (due to the codewords associate with unique combination of read sequences that allow error-detection step), as opposed to being mis-detected as was previously commonplace. Thus, the determination of codewords and matching them to the valid codewords associated with the nucleic acid target, when considered as a combination with the other additional elements, yields a claim as a whole that amounts to

significantly more than the exception itself. *See*, e.g., *Finjan Inc. v. Blue Coat Systems, Inc.*, 879 F.3d 1299 (Fed. Cir. 2018) (claims to virus scanning that can protect the user against both previously unknown viruses and obfuscated code were found to be an improvement over traditional virus scanning); *see also Diamond v. Diehr*, 450 U.S. 175, 188 (1981) ("a new combination of steps in a process may be patentable even though all the constituents of the combination were well known and in common use before the combination was made"). Accordingly, claim 1 and its dependent claims are patent eligible, and withdrawal of this rejection is respectfully requested.

CONCLUSION

In view of the foregoing remarks, this application should now be in condition for allowance. A notice to this effect is respectfully requested. If the Examiner believes, after the foregoing remarks, that the application is not in condition for allowance, the Examiner is requested to call the undersigned at 617-646-8000.

If this response is not considered timely filed and if a suitable request for an extension of time is otherwise absent, any necessary extension of time is hereby requested. If there is a fee occasioned by this response, including an extension fee, the Director is hereby authorized to charge any deficiency or credit any overpayment in the fees filed, asserted to be filed or which should have been filed herewith to Deposit Account No. 23/2825, under Docket No. H0498.70509US02 from which the undersigned is authorized to draw.

Dated: January 22, 2021

Respectfully submitted,

Electronic signature: /Tani Chen/
Tani Chen, Sc.D.
Registration No.: 52,728
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

8660010.1

IN THE CLAIMS

Please replace all prior versions, and listings, of claims in the application with the following list of claims. Additions are indicated by underlining and deletions are indicate by strikeouts and/or double bracketing.

1. (Currently amended) A method for detecting nucleic acid targets, comprising:

~~(a) associating a plurality of nucleic acid targets with valid codewords, comprising:~~

~~(i) defining a codeword space comprising a plurality of valid codewords, wherein the valid codewords in the codeword space are defined by a series of positions and values at the positions, and the valid codewords form an error-checking and/or error-correcting code space; and~~

~~(ii) associating the valid codewords with a plurality of nucleic acids targets such that each nucleic acid target is associated with a unique valid codeword;~~

~~(b)~~ (a) exposing a sample suspected of containing at least some of ~~the~~ a plurality of nucleic acid targets to a plurality of encoding nucleic acid probes ~~that contain a representation of the valid codewords, comprising:~~.

wherein at least some of the plurality of encoding nucleic acid probes comprise a target sequence that specifically hybridizes to a nucleic acid target in the plurality of nucleic acid targets and one or more read sequences,

wherein unique combinations of one or more read sequences correspond to valid codewords, each of the valid codewords associated to each of the plurality of nucleic acid targets, wherein the valid codeword consists of a series of positions and values for each position, and

~~(i)~~ ~~providing a plurality of nucleic acid read sequences and associating a unique~~ each read sequence is associated with each of the values, or a subset thereof, at each of the positions of the valid codewords;

~~(ii) creating a plurality of encoding nucleic acid probes, wherein each of the encoding nucleic acid probes comprises a first portion comprising a target sequence able to bind to one of the nucleic acid targets via nucleic acid hybridization, and a second portion~~

~~comprising one or more read sequences taken from the plurality of read sequences; and~~

~~(iii) exposing the sample to the plurality of encoding nucleic acid probes to cause the encoding nucleic acid probes to bind to at least some of the nucleic acid targets in the sample via nucleic acid hybridization; and~~

~~(c) identifying nucleic acid targets in the sample by determining valid codewords that correspond to the plurality of nucleic acid probes, comprising:~~

~~(i)~~ (b) exposing the sample to a first round of one or more readout nucleic acid probes, wherein at least some of the readout nucleic acid probes comprise a first portion and a second portion, the first portion comprising a sequence ~~substantially complementary~~ that specifically hybridizes to one of the read sequences, and the second portion comprising a fluorescent signaling entity;

~~(ii) imaging the sample to determine a fluorescence signal of the fluorescent signaling entity in at least some locations in the sample;~~

(c) identifying nucleic acid targets in the sample, comprising:

detecting fluorescent signals emitted from the fluorescent signaling entities of bound readout nucleic acid probes in at least some locations within the sample to determine a codeword in at least some locations within the sample based on the read sequences present on the bound encoding nucleic acid probes;

identifying the codeword at each location by comparing the codeword to the valid codewords associated with the plurality of nucleic acid targets, and if no match is found, applying error correction to the codeword to match one of the valid codewords or discarding the codeword; and

assigning the identity of some of the plurality of nucleic acid targets in the sample based on the identity of the nucleic acid targets associated with the valid codewords to which the identified codewords were matched

~~(iii) determining read sequences and codewords at the locations based on the fluorescence signal;~~

~~(iv) matching the codewords tothe valid codewords in the codeword space wherein, if no match is found, either discarding the codeword or applying error correction to~~

~~the codeword to match one of the valid codewords; and~~

~~(v) identifying nucleic acid targets in the sample based on the codewords matched to the valid codewords~~.

2-33. (Cancelled)

34. (Currently amended) The method of claim 1, further comprising inactivating at least some of the fluorescent signaling entity and/or removing at least some of the fluorescent signaling entity from the sample after said exposing the sample to the one or more readout nucleic acid probes and imaging the sample to determine the fluorescence signal of the fluorescent signaling entities.

35-70. (Cancelled)

71. (Previously presented) The method of claim 1, comprising determining the fluorescent signaling entity at a resolution better than 300 nm.

72-74. (Cancelled)

75. (Previously presented) The method of claim 1, wherein the sample comprises a cell.

76. (Cancelled)

77. (Previously presented) The method of claim 75, wherein the cell is fixed.

78. (Previously presented) The method of claim 1, comprising determining the fluorescent signal entity by imaging at least a portion of the sample using a light microscope.

79-176. (Cancelled)

177. (Currently amended) The method of claim 1, further comprising:

exposing the sample to a second round of one or more readout nucleic acid probes, wherein at least some of the readout nucleic acid probes of the second round comprise a first portion and a second portion, the first portion comprising a sequence ~~substantially complementary~~ that specifically hybridizes to one of the read sequences different from the read sequences of the first round, and the second portion comprising a fluorescent signaling entity; and

imaging the sample to determine a fluorescence signal of the fluorescent signaling entity of the second round of one or more readout nucleic acid probes in at least some locations in the sample.

178. (Currently amended) The method of claim 1, wherein the ~~codeword position~~ values at each position in the codewords are binary values.

179. (Currently amended) The method of claim 1, further comprising determining an abundance and/or spatial distribution of the nucleic acid targets based on the codewords matched to the valid codewords in each location within the sample.

180. (Currently amended) The method of claim 1, further comprising removing unbound readout nucleic acid probes prior to said imaging the sample.

181. (Currently amended) The method of claim 34, wherein ~~the acts of~~ said exposing the sample to one or more readout nucleic acid probes, said imaging the sample, and said inactivating at least some of the fluorescent signaling entities on the read out probes are performed ~~for~~ to assign each of the values, or subset thereof, at each of the positions within the ~~codeword space~~ codewords.

182. (Cancelled)

183. (New) The method of claim 1, wherein the valid codewords form an error-checking or an error-correcting code space.

184. (New) The method of claim 1, wherein in comparing the codeword to the valid codewords in step (c), no match is found when the codeword differs from the valid codeword compared by one or more errors.

185. (New) The method of claim 1, wherein in comparing the codeword to the valid codewords in step (c), no match is found when the codeword differs from the valid codeword compared by one error.

186. (New) The method of claim 1, wherein in comparing the codeword to the valid codewords in step (c), no match is found when the codeword differs from the valid codeword compared by two errors.

187. (New) The method of claim 1, wherein comparing the codeword to the valid codewords in step (c) comprises the values at each of the positions of the codewords.

FORM PTO-1449/A and B (modified PTO/SB/08)				APPLICATION NO.: 15/329,683	ATTY. DOCKET NO.: H0498.70509US02
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				FILING DATE: January 27, 2017	CONFIRMATION NO.: 7428
				FIRST NAMED INVENTOR: Xiaowei Zhuang	
Sheet	1	of	2	GROUP ART UNIT: 1634	EXAMINER: F. W. Lu

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document		Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		Number	Kind Code		
		2015-144490	A1	Deisseroth et al.	05-28-2015
		2016-0153005	A1	Cong et al.	06-02-2016
		2018-0230226	A1	Podack et al.	08-16-2018

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document			Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)
		Office/ Country	Number	Kind Code			
		WO	2018/089438	A1	President and Fellows of Harvard College	05-17-2018	

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		European Office Action mailed October 16, 2020 for Application No. 15827358.1 (H0498.70509EP00).	
		Chinese Office Action dated August 19, 2020 for Application No. CN 201580051029.5 (H0498.70510CN00).	Y - Full
		European Office Action mailed August 19, 2020 for Application No. 15828133.7 (H0498.70510EP00).	
		Extended European Search Report mailed May 20, 2020 for Application No. 17869122.6 (H0498.70630EP00).	
		Supplementary European Search Report mailed June 9, 2020 for Application No. 17869122.6 (H0498.70630EP00).	
		Extended European Search Report mailed July 20, 2020 for Application No. 17869515.1 (H0498.70645EP00).	
		European Office Action mailed August 6, 2020 for Application No. 17869515.1 (H0498.70645EP00).	
		International Search Report and Written Opinion for Application No. PCT/US2019/065857 mailed May 1, 2020 (H0498.70680WO00).	
		International Search Report and Written Opinion for Application No. PCT/US20/28632 mailed July 16, 2020 (H0498.70693WO00).	
		CHEN et al., RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science. 2015;348(6233):aaa6090.	
		CHEN et al., Supplemental Material for Expansion microscopy. Science. 2015 Jan; 347(6221):1-18.	
		CHUNG et al., CLARITY for mapping the nervous system [published correction appears in Nat Methods. 2013 Oct;10(10):1035]. Nat Methods. 2013;10(6):508-513.	
		TSANOV et al., smiFISH and FISH-quant - a flexible single RNA detection approach with super-resolution capability. Nucleic Acids Res. 2016;44(22):e165.	
		YANG et al., Single-cell phenotyping within transparent intact tissue through whole-body clearing. Cell. 2014;158(4):945-958.	

EXAMINER:	DATE CONSIDERED:

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

8892647.1

FORM PTO-1449/A and B (modified PTO/SB/08)				APPLICATION NO.: 15/329,683	ATTY. DOCKET NO.: H0498.70509US02
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				FILING DATE: January 27, 2017	CONFIRMATION NO.: 7428
				FIRST NAMED INVENTOR: Xiaowei Zhuang	
Sheet	2	of	2	GROUP ART UNIT: 1634	EXAMINER: F. W. Lu

		ZHANG et al., Tandem Spinach Array for mRNA Imaging in Living Bacterial Cells. Sci Rep. 2015;5:17295.	

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

EXAMINER:	DATE CONSIDERED:

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

8892647.1

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 May 2018 (17.05.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/089438 A1

(51) International Patent Classification:
G01N 1/36 (2006.01) *C12Q 1/68* (2018.01)

(21) International Application Number:
PCT/US2017/060558

(22) International Filing Date:
08 November 2017 (08.11.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/419,033 08 November 2016 (08.11.2016) US
62/569,127 06 October 2017 (06.10.2017) US

(71) Applicant: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).

(72) Inventors: **MOFFITT, Jeffrey, R.**; 175 Beacon Street, Apt. #108, Somerville, MA 02143 (US). **WANG, Guiping**; 57 Springfield St. Apt. 3, Somerville, MA 02143 (US). **ZHUANG, Xiaowei**; 5 Luongo Farm Lane, Lexington, MA 02421 (US).

(74) Agent: **CHEN, Tani**; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*

(54) Title: MULTIPLEXED IMAGING USING MERFISH, EXPANSION MICROSCOPY, AND RELATED TECHNOLOGIES

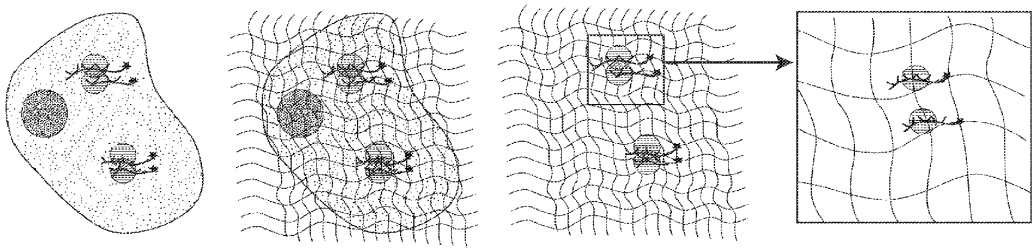


FIG. 2A

(57) Abstract: The present invention generally relates to microscopy, and to systems and methods for imaging or determining nucleic acids or other desired targets, for instance, within cells. In certain aspects, a sample is contained within an expandable material, which is expanded and imaged in some fashion. Expansion of the material improves the effective resolution of the subsequent image. This may be combined, for example, with other super-resolution techniques, such as STORM, and/or with techniques such as MERFISH for determining nucleic acids such as mRNA within the sample, for example, by binding nucleic acid probes to the sample. Other aspects are generally directed to compositions or devices for use in such methods, kits for use in such methods, or the like.

WO 2018/089438 A1

MULTIPLEXED IMAGING USING MERFISH, EXPANSION MICROSCOPY, AND RELATED TECHNOLOGIES

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Patent Application Serial No. 62/419,033, filed November 8, 2016, entitled "Matrix Imprinting and Clearing," by Zhuang, *et al.*, and U.S. Provisional Patent Application Serial No. 62/569,127, filed October 6, 2017, entitled "Multiplexed Imaging Using MERFISH, Expansion Microscopy, and Related Technologies," by Zhuang, *et al.* Each of these is incorporated herein by reference in its entirety.

GOVERNMENT FUNDING

This invention was made with government support under Grant. No. OD022125 awarded by the National Institutes of Health. The government has certain rights in the invention.

FIELD

The present invention generally relates to microscopy, and to systems and methods for imaging or determining nucleic acids or other desired targets, for instance, within cells.

BACKGROUND

Image-based single-cell transcriptomics, in which RNA species are identified, counted and localized *in situ* via imaging, naturally preserve the native spatial context of RNAs. Multiplexed error-robust fluorescence *in situ* hybridization (MERFISH) is a method for single-cell transcriptome imaging and has been demonstrated to profile hundreds to thousands of RNAs in single cells. In MERFISH, RNAs are identified via a combinatorial labeling approach that encodes RNA species with error-robust binary barcodes followed by sequential rounds of single-molecule fluorescence *in situ* hybridization (smFISH) to read out these barcodes. The accuracy of the RNA identification relies on spatially separated signals of individual RNA molecules, which limits the density of RNAs that can be measured and makes the multiplexed imaging of a large number of high abundance RNAs challenging. Accordingly, improvements in imaging techniques are needed.

SUMMARY

The present invention generally relates to microscopy, and to systems and methods for imaging or determining nucleic acids or other desired targets, for instance, within cells. The subject matter of the present invention involves, in some cases, interrelated products,

alternative solutions to a particular problem, and/or a plurality of different uses of one or more systems and/or articles.

In one aspect, the present invention is generally directed to an article comprising an expandable material. The material may comprise an embedded cell and a plurality of nucleic acids immobilized to the expandable material. In some cases, at least 50% of the plurality of nucleic acids immobilized to the expandable material are immobilized at single points.

According to another aspect, the present invention is generally directed to a polymer comprising an expanded cell and a plurality of nucleic acids immobilized to the polymer. In some embodiments, the cell is expanded to at least 5 times its normal size within the polymer. In certain instances, at least 50% of the plurality of nucleic acids immobilized to the expandable material are immobilized at single points.

The present invention, in another aspect, is generally directed to a method. In one set of embodiments, the method comprises embedding cells within an expandable material, immobilizing nucleic acids from the cells to the expandable material, expanding the expandable material, exposing the expandable material to a plurality of nucleic acid probes, and determining binding of the nucleic acid probes to the immobilized nucleic acids.

In another set of embodiments, the method comprises immobilizing a plurality of targets to an expandable material, and expanding the expandable material. In some cases, at least 50% of the plurality of targets immobilized to the expandable material are immobilized at single points.

The method, in yet another set of embodiments, includes immobilizing a plurality of nucleic acids to an expandable material, exposing the expandable material to a plurality of nucleic acid probes, expanding the expandable material, and determining binding of the nucleic acid probes to the immobilized nucleic acids. In certain embodiments, at least 50% of the plurality of nucleic acids immobilized to the expandable material are immobilized at single points.

In another aspect, the present invention encompasses methods of making one or more of the embodiments described herein, for example, MERFISH and expansion microscopy. In still another aspect, the present invention encompasses methods of using one or more of the embodiments described herein, for example, MERFISH and expansion microscopy.

Other advantages and novel features of the present invention will become apparent from the following detailed description of various non-limiting embodiments of the invention when considered in conjunction with the accompanying figures.

BRIEF DESCRIPTION OF THE DRAWINGS

Non-limiting embodiments of the present invention will be described by way of example with reference to the accompanying figures, which are schematic and are not intended to be drawn to scale. In the figures, each identical or nearly identical component illustrated is typically represented by a single numeral. For purposes of clarity, not every component is labeled in every figure, nor is every component of each embodiment of the invention shown where illustration is not necessary to allow those of ordinary skill in the art to understand the invention. In the figures:

Figs. 1A-1E illustrate MERFISH measurements of RNA in an unexpanded sample;

Figs. 2A-2H illustrate MERFISH measurements of RNA in an expanded sample, in one embodiment of the invention;

Figs. 3A-3H illustrate imaging in expanded samples in combination with MERFISH, in yet another embodiment of the invention; and

Figs. 4A-4C schematically illustrate attachment of targets at one or more anchor points.

DETAILED DESCRIPTION

The present invention generally relates to microscopy, and to systems and methods for imaging or determining nucleic acids or other desired targets, for instance, within cells. In certain aspects, a sample is contained within an expandable material, which is expanded and imaged in some fashion. Expansion of the material improves the effective resolution of the subsequent image. This may be combined, for example, with other super-resolution techniques, such as STORM, and/or with techniques such as MERFISH for determining nucleic acids such as mRNA within the sample, for example, by binding nucleic acid probes to the sample. Other aspects are generally directed to compositions or devices for use in such methods, kits for use in such methods, or the like.

In certain aspects, a sample is embedded or contained within an expandable material, such as polyelectrolyte gel. The sample may be, for example, a cell or other biological structure, or a non-biological structure in some cases. Upon expansion of the expandable material, the sample is effectively magnified physically, rather than optically. Components of the sample are separated from each other by expansion of the material, but typically retain comparable geometries or topologies, especially if the expansion of the material occurs substantially isotopically. Thus, components that initially started close to each other are now further apart, and can be more easily detected, e.g., using imaging or other techniques.

Expansion techniques may be combined with techniques for determining nucleic acids such as mRNA within the sample. In some cases, for example, nucleic acids within the

sample are attached to the expandable material (for example, covalently) before the expandable material is expanded. The nucleic acids may then be detected, for example, using exposures to nucleic acid probes, e.g., repeatedly, which may be used to determine the distribution of nucleic acids within the material. For example, in techniques such as MERFISH (multiplexed error-robust fluorescence *in situ* hybridization), a sample is exposed to different rounds of nucleic acid probes, and binding of the nucleic acids can be determined using fluorescence or other techniques. In certain cases, a relatively large number of different targets may be identified using a relatively small number of labels, e.g., by using various combinatorial approaches. Identification may also be enhanced in some embodiments using error-checking and/or error-correcting codes. See, e.g., U.S. Patent Application Serial No. 15/329,683, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*, published as U.S. Patent Application Publication No. 2017/0220733 on August 3, 2017, incorporated herein by reference in its entirety.

In some cases, one or more targets, for example, nucleic acids such as mRNA, may be immobilized relative to the expandable material prior to expansion. In certain embodiments, targets are attached to the expandable material at a single point. This is important in some embodiments since targets attached at two or more points may be expanded along with the material, which may make it more difficult to resolve targets at a high resolution. For example, referring to Fig. 4A, targets 10 and 11 are shown within an expandable material 20. For example, targets 10 and 11 may be nucleic acids. In Fig. 4B, targets 10 and 11 are attached at two points (e.g., their ends), and upon expansion by the expandable material, targets 10 and 11 are expanded (e.g., "stretched") along with the rest of the expanded material since their anchor points (shown as dots) are expanded along with the expanded material. In contrast, in Fig. 4C, targets 10 and 11 are attached at only one point (shown as a dot), and upon expansion of the expandable material, the targets are not expanded or stretched. Thus, in Fig. 4C, targets 10 and 11 are now further apart, and can be more easily determined, whereas in Fig. 4B targets 10 and 11 may still be physically overlapping and, thus, may produce signals that are not distinguishable despite the expansion process. Accordingly, in certain embodiments of the invention, at least some of the targets, such as nucleic acids, are attached to the expandable material at a single point. In some embodiments, a molecule may behave in the same way as described in Fig. 4C even if it is attached to the gel via multiple attachments, e.g., if the attachments are sufficiently close to one another along the length of the molecule.

In addition, components of the sample that may be contributing to the background, such as proteins, lipids, and/or other non-targets, may be "cleared" from the sample to improve determination in certain cases. However, nucleic acids or other desired targets may be prevented from also being cleared, for example, by attaching the nucleic acids (or other targets) to the expandable material, as noted above. In this way, expandable materials may be combined with techniques such as MERFISH to facilitate detection of mRNA, other nucleic acids, or other targets. Non-limiting examples of such techniques may be found, for example, in U.S. Pat. Apl. Pub. No. 62/419,033, entitled "Matrix Imprinting and Clearing," incorporated herein by reference in its entirety.

A variety of techniques may be used to determine binding, including optical techniques such as fluorescence microscopy. In some cases, spatial positions may be determined at super resolutions, or at resolutions better than the wavelength of light or the diffraction limit (although in other embodiments, super resolutions are not required). For example, techniques such as STORM (stochastic optical reconstruction microscopy) may be used. See, for example, U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*, incorporated herein by reference in its entirety.

The above discussion is a non-limiting example of one embodiment of the present invention that can be used to determine nucleic acids or other targets in a sample. However, other embodiments are also possible. Accordingly, more generally, various aspects of the invention are directed to various systems and methods for imaging or determining nucleic acids or other desired targets, for instance, within cells or other samples.

As mentioned, in one aspect, a sample is embedded or contained within an expandable material. The sample may be any suitable sample, and may be biological in some embodiments. In some cases, the sample contains DNA and/or RNA, e.g., that may be determined within the sample. (In other embodiments, other targets within the sample may be determined.) In some cases, the sample may include cells, such as mammalian cells (including human cells), or other types of cells. The sample may contain viruses in some cases. In addition, in some cases, the sample may be a tissue sample, e.g., from a biopsy, artificially grown or cultured, etc.

In certain embodiments, the expandable material is one that can be expanded, for example, when exposed to water or another suitable liquid. For example, the material may exhibit a relative change in size of at least 1.1, at least 1.2 at least 1.3, at least 1.5, at least 2, at least 3, at least 4, at least 5, at least 7, at least 10, or at least 15, etc., and/or a relative

change in size that is less than 15, less than 10, less than 7, less than 5, less than 4, less than 3, less than 2, less than 1.5, less than 1.3, or less than 1.2 (i.e., a change in size of 2 means that a sample doubles in linear dimension), or inverses of these (i.e., an inverse change in size of 2 means that a sample halves in linear dimensions).

In some embodiments, the expandable material may be one that does not significantly distort during the expansion process (e.g., the expandable material may expand substantially uniformly or isotropically in all 3 dimensions), although in some cases, the expandable material may exhibit some distortion or non-isotropic expansion. For example, the expandable material may expand in one dimension, relative to an orthogonal dimension, by less than 150%, less than 130%, less than 125%, less than 120%, less than 115%, less than 110%, or less than 105% by linear dimension relative to the shorter linear expansion.

In some cases, the expandable material is a polymer. Non-limiting examples of suitable polymers include polyelectrolytes and agarose. In some cases, the polymer is a gel or a hydrogel. A variety of polymers can be used in various embodiments including but not limited to acrylic acid, acrylamide, ethylene glycol diacrylate, ethylene glycol dimetharcrylate, poly(ethylene glycol dimethacrylate), poly(*N*-isopropyl acrylamide), methyl cellulose, (ethylene oxide)-(propylene oxide)-(ethylene oxide) terpolymers, sodium alginate, poly(vinyl alcohol), alginate, chitosan, gum Arabic, gelatin, agarose, or the like. In some cases, the polymer may be selected to be relatively optically transparent. In some cases, the expandable material may be formed from monomers or oligomers, for example, comprising one or more substituted or unsubstituted methacrylates, acrylates, acrylamides, methacrylamides, vinylalcohols, vinylamines, allylamines, allylalcohols, including divinylic crosslinkers thereof (e.g., *N*,*N*-alkylene bisacrylamides such as *N*,*N*-methylenebisacrylamide), or the like. In some cases, polymerization initiators and/or crosslinkers may be present. For example, a precursor may include one or more cross-linking agents, which may be used to cross-link a polymeric expandable material as it forms, e.g., during the polymerization process.

In some cases, expansion of the expandable material may be facilitated by exposing the material to water, a solution comprising water, or another suitable medium (e.g., a liquid medium). Without wishing to be bound by any theory, it is believed that water flow into the material may facilitate the expansion of the expandable material. In certain embodiments, the solution may be hypotonic relative to the expandable material, which may facilitate the transport of water into the expandable material, e.g., due to differences in tonicity. Other ways of expanding polymers may be used in other embodiments, such as through changes in

the pH, changes in an external electric/magnetic field, changes in temperature, response to light, etc.

In some cases, expansion of the material may be restricted after the expansion has occurred. For example, the material may be stabilized so that exchange of other buffers does not cause a change in size. This stabilization may occur, for example, via chemical modification of the expandable material, or embedding of the expandable material in a second material that is not expandable.

As a non-limiting example, a sample may be exposed to one or more monomers or other precursors which can react to form the expandable material. In some cases, the precursors may be permeated, diffused, or otherwise transported through the sample, before being reacted. For instance, in some cases, a precursor may be present in a liquid (e.g., water, saline, or other aqueous medium), which may permeate through a sample in some fashion.

In some cases, the sample may be embedded within a relatively large polymer or gel, which can then be sectioned or sliced in some cases to produce smaller portions for analysis, e.g., using various microtomy techniques commonly available to those of ordinary skill in the art. For instance, tissues or organs may be immobilized within a suitable polymer or gel.

In certain embodiments, the expansion of the gel may disentangle molecules that were physically overlapping. In a non-limiting example, two RNA molecules may physically overlap within the sample. By anchoring these RNA molecules at one location, the expansion of the gel may separate these molecules so that they are no longer physically overlapping as long as the attachment points for each RNA are sufficiently separated within the sample. In some embodiments, for example, if the RNA molecules are anchored to the gel at multiple locations along their length, then expansion of the gel may not physically separate these two molecules. Rather, expansion of the gel may stretch these molecules between the attachment points, leaving the two regions of these molecules that were overlapping still overlapping after expansion. In this case, these molecules may not be capable of being physically distinguished, e.g., by optical techniques. In some cases, anchoring methods that are deterministic (e.g., that target a specific location of the RNA) versus random (e.g., that target a large number of different locations on the RNA with the specific targeted location selected at random), may differ in the probability that RNAs will be anchored in multiple locations to the gel and, thus, be unable to be physically separated when overlapping. Specific targeting of the 3' or 5' end, or a defined location, within the RNA may in some cases lead to deterministic anchoring that will not stretch RNAs. Moreover, it should be understood that

although RNA was used in this example, such considerations also apply to any other biomolecules within samples, e.g. DNA, or other targets described herein.

In one aspect, a sample that is expanded may be imaged or studied to determine nucleic acids or other desired targets, for instance, within cells, tissues or other samples contained within an expandable material. Techniques useful for determining nucleic acids or other desired targets include, but are not limited to, MERFISH, smFISH, or techniques such as those disclosed in U.S. Patent Application Serial No. 15/329,683, filed January 27, 2017, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*, published as U.S. Patent Application Publication No. 2017/0220733 on August 3, 2017; or U.S. Patent Application Serial No. 15/329,651, filed January 27, 2017, entitled "Probe Library Construction," by Zhuang, *et al.*, published as U.S. Patent Application Publication No. 2017/0212986 on July 27, 2017; each incorporated herein by reference in its entirety. In addition, in some cases, a desired target may be immobilized within the expandable material (such as a polymer or gel), while other components are "cleared," e.g., via degradation and/or physical removal, for example, as discussed in U.S. Provisional Patent Application Serial No. 62/419,033, filed November 8, 2016, entitled "Matrix Imprinting and Clearing," by Zhuang, *et al.*, incorporated herein by reference in its entirety.

If nucleic acids are desired to be determined, the nucleic acids may be, for example, DNA, RNA, or other nucleic acids that are present within a cell (or other sample). The nucleic acids may be endogenous to the cell, or added to the cell. For instance, the nucleic acid may be viral, or artificially created. In some cases, the nucleic acid to be determined may be expressed by the cell. The nucleic acid is RNA in some embodiments. The RNA may be coding and/or non-coding RNA. Non-limiting examples of RNA that may be studied within the cell include mRNA, siRNA, rRNA, miRNA, tRNA, lncRNA, snoRNAs, snRNAs, exRNAs, piRNAs, or the like.

In some cases, a significant portion of the nucleic acid within the cell may be studied. For instance, in some cases, enough of the RNA present within a cell may be determined so as to produce a partial or complete transcriptome of the cell. In some cases, at least 4 types of mRNAs are determined within a cell, and in some cases, at least 3, at least 4, at least 7, at least 8, at least 12, at least 14, at least 15, at least 16, at least 22, at least 30, at least 31, at least 32, at least 50, at least 63, at least 64, at least 72, at least 75, at least 100, at least 127, at least 128, at least 140, at least 255, at least 256, at least 500, at least 1,000, at least 1,500, at least 2,000, at least 2,500, at least 3,000, at least 4,000, at least 5,000, at least 7,500, at least 10,000, at least 12,000, at least 15,000, at least 20,000, at least 25,000, at least 30,000, at least

40,000, at least 50,000, at least 75,000, or at least 100,000 types of mRNAs may be determined within a cell.

In some cases, the transcriptome of a cell may be determined. It should be understood that the transcriptome generally encompasses all RNA molecules produced within a cell, not just mRNA. Thus, for instance, the transcriptome may also include rRNA, tRNA, siRNA, etc. In some embodiments, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or 100% of the transcriptome of a cell may be determined.

The determination of one or more nucleic acids within the cell or other sample may be qualitative and/or quantitative. In addition, the determination may also be spatial, e.g., the position of the nucleic acid within the cell or other sample may be determined in two or three dimensions. In some embodiments, the positions, number, and/or concentrations of nucleic acids within the cell (or other sample) may be determined.

In some cases, a significant portion of the genome of a cell may be determined. The determined genomic segments may be continuous or interspersed on the genome. For example, in some cases, at least 4 genomic segments are determined within a cell, and in some cases, at least 3, at least 4, at least 7, at least 8, at least 12, at least 14, at least 15, at least 16, at least 22, at least 30, at least 31, at least 32, at least 50, at least 63, at least 64, at least 72, at least 75, at least 100, at least 127, at least 128, at least 140, at least 255, at least 256, at least 500, at least 1,000, at least 1,500, at least 2,000, at least 2,500, at least 3,000, at least 4,000, at least 5,000, at least 7,500, at least 10,000, at least 12,000, at least 15,000, at least 20,000, at least 25,000, at least 30,000, at least 40,000, at least 50,000, at least 75,000, or at least 100,000 genomic segments may be determined within a cell.

In some cases, the entire genome of a cell may be determined. It should be understood that the genome generally encompasses all DNA molecules produced within a cell, not just chromosome DNA. Thus, for instance, the genome may also include, in some cases, mitochondria DNA, chloroplast DNA, plasmid DNA, etc. In some embodiments, at least about 5%, at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or 100% of the genome of a cell may be determined.

However, as discussed, it should be understood that in other embodiments of the invention, other targets may be determined or immobilized, e.g., in addition to and/or instead of nucleic acids. For example, in some embodiments of the invention, the targets to be

determined or immobilized may include proteins (e.g., antibodies, enzymes, structural proteins), lipids, carbohydrates, viruses, or the like. In one embodiment, cellular components, such as proteins, can be detected by binding to them proteins, such as antibodies/immunoglobulins (primary antibodies, secondary antibodies, nanobodies, fragments of antibodies, IgG, IgM, IgA, IgD, IgE, etc.), that are conjugated to oligonucleotide probes which are anchored to the polymer or gel. These components could then be removed, leaving the oligonucleotide probes to be detected via hybridization of additional nucleic acid probes, similar or identical to the detection of cellular nucleic acids. In another embodiment, multiple distinct cellular species could be detected simultaneously within the same sample, even if the original components are removed from the gel or polymer. For example, RNA molecules could be detected via hybridization of nucleic acid probes simultaneously with the detection of proteins via antibody-oligonucleotide conjugates, as described above.

As mentioned, the sample can be immobilized or embedded within a polymer or a gel, partially or completely. For example, the sample may be embedded in an expandable material as discussed above. In one set of embodiments, anchor probes may be used during the polymerization process. The anchor probes may include an anchor portion that is able to polymerize with the expandable material, e.g., during and/or after the polymerization process, and a targeting portion that is able to immobilize a target, e.g., chemically and/or physically. For example, in the case of polyacrylamide, the anchor probe may include an acrydite portion that can polymerize and become incorporated into the polymer. As another example, an anchor probe may contain, as a targeting portion, a sequence of nucleic acids that is complementary to a target that is a nucleic acid, such as RNA (e.g., mRNA) or DNA. The targeting portion may be specific to a target, and/or may randomly associate with different targets within a sample (for example, due to non-specific binding). Other portions may be present within the anchor probes as well.

For example, to associate with a target nucleic acid, the anchor probe may comprise a nucleic acid sequence substantially complementary to at least a portion of the target nucleic acid. For instance, the nucleic acid may be complementary to at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 or more nucleotides of the nucleic acid. In some cases the complementarity may be exact (Watson-Crick complementarity), or there may be 1, 2, or more mismatches.

Thus, the anchor probe may contain a portion that can interact with and bind to nucleic acid molecules in some embodiments, and/or other molecules in which

immobilization is desired, e.g., proteins or lipids, other desired targets, etc. The immobilization may be covalent or non-covalent. For example, to immobilize a target nucleic acid, the anchor probe may comprise a nucleic acid comprising an acrydite portion (e.g., at the 5' end, the 3' end, an internal base, etc.), and a portion able to recognize the target nucleic acid.

In some cases, the anchor probe can be configured to immobilize mRNA, e.g., in the case of transcriptome analysis. For instance, in one set of embodiments, the anchor probe may contain a plurality of thymine nucleotides, e.g., sequentially, for binding to the poly-A tail of an mRNA. Thus, for example, the anchor probe can have at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 or more consecutive thymine nucleotides (e.g., a poly-dT portion) within the anchor probe. In some cases, at least some of the thymine nucleotides may be "locked" thymine nucleotides. These may comprise at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, or at least 80% of these thymine nucleotides. In certain embodiments, the locked and non-locked nucleotides may alternate. Such locked thymine nucleotides may be useful, for example, to stabilize the hybridization of the poly-A tails of the mRNA with the anchor probe.

In another set of embodiments, the anchor probe may comprise a sequence substantially complementary to mRNA (or another target nucleic acid), as noted above. The sequence may be substantially complementary to all, or only a portion, of the target nucleic acid, for example, an end portion (e.g., towards a 5' end or a 3' end), or a middle portion between the end portions. For example, a nucleic acid may be immobilized using anchor probes having substantially complementary portions to the DNA or RNA target. There may be, e.g., 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 12 or more, 13 or more, 14 or more, 15 or more, 20 or more, 25 or more, 30 or more, 35 or more, 40 or more, 45 or more, or 50 or more complementary nucleotides between the anchor probe and the nucleic acid.

In another set of embodiments, a plurality of anchor probes may be targeted via the methods described above to each RNA of interests. For example, 2 or more, 3 or more, 4 or more, 5 or more, 10 or more, 15 or more, or 20 or more distinct nucleic acid anchor probes may be targeted to a given RNA. These probes may be targeted to a small region of this RNA such that if multiple probes bind and anchor the RNA to the gel, the majority of the RNA remains unstretched during expansion. If multiple RNA species are targeted at a time, then the number of unique anchor probes applied to the sample may be, for example, 5 or more, 10

or more, 15 or more, 20 or more, 100 or more, 200 or more, 1,000 or more, or 10,000 or more.

Other methods may be used to anchor nucleic acids, or other molecules in which immobilization is desired. In one set of embodiments, nucleic acids such as DNA or RNA may be immobilized by covalent bonding. For example, in one set of embodiments, an alkylating agent may be used that covalently binds to RNA or DNA and contains a second chemical moiety that can be incorporated into the polyelectrolytes as it is polymerized. In yet another set of embodiments, the terminal ribose in an RNA molecule may be oxidized using sodium periodate (or another oxidizing agent) to produce an aldehyde, which may be cross-linked to acrylamide, or other polymer or gel. In other embodiments, chemical agents that are able to modify bases may be used, such as aldehydes, e.g. paraformaldehyde or gluteraldehyde, alkylating agents, or succinimidyl-containing groups; chemical agents that modify the terminal phosphate, such as carbodiimides, e.g., EDC (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide); chemical agents that modify internal sugars, such as p-maleimido-phenyl isocyanate; or chemical agents that modify terminal sugars, such as sodium periodate. In some cases, these chemical agents can carry a second chemical moiety that can then be directly cross-linked to the gel or polymer, and/or which can be further modified with a compound that can be directly cross linked to the gel or polymer.

In still another set of embodiments, the nucleic acids may be physically tangled within the polymer or gel, e.g., due to their length, and, thus, unable to diffuse from their original location within the gel.

Similar anchor probes may be used to immobilize other components to a polymer or gel, in other embodiments. For example, in one set of embodiments, an antibody able to specifically bind to a suitable target (e.g., another protein, a lipid, a carbohydrate, a virus, etc.) may be modified to include an acrydite moiety that can become incorporated within a polymer or gel.

In addition, it should be understood that the embedding of the sample within the expandable material and the immobilization of nucleic acids (or other desired targets) may be performed in any suitable order in various embodiments. For instance, immobilization may occur before, during, or after embedding of the sample. In some cases, the target may be chemically modified or reacted to cross-link to the gel or polymer before or during formation of the gel or polymer.

As mentioned, in some embodiments of the invention, the anchor probes immobilize a target, such as a nucleic acid, to the expandable material at a single point. In some cases, at

least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 95% (by number) of the targets immobilized to the expandable material are immobilized at a single point. By using only a single point, relatively large numbers of targets may be immobilized relative to the expandable material, without necessarily affecting function of the targets. For example, multiple points of attachment may potentially disrupt the structure of the target, e.g., due to "stretching" of the target during expansion of the expandable material (see, e.g., Figs. 4A-4C), or due to substitutions or distortions of the target due to the anchor (for example, if the anchoring occurs at or near a site used for binding or structural stability, etc.). However, it should be understood that in some embodiments, targets may be attached to the expandable material at more than one point.

A variety of techniques may be used to attach a target such as a nucleic acid (e.g., RNA or DNA) at a single point. For example, anchor probes may be used that are complementary to specific regions of a nucleic acid (e.g., based on sequence complementarity). In some cases, the anchor probes may be specific to each nucleic acid individually (e.g., having unique nucleic acid sequences that are complementary to that nucleic acid's sequence). Nucleic acids such as DNA or RNA can thus be immobilized to an expandable material, e.g., at a single point or region controlled by the sequence of nucleic acids within the anchor probe. In other embodiments, however, the anchor probes may include nucleic acids that are complementary to a common feature of the nucleic acid, such as the poly-A tail of an mRNA, such that the same anchor probes can immobilize several different nucleic acids (e.g., different mRNA sequences within a sample).

In addition, in some embodiments, more than one type of anchor probe may be used, for example, targeting the same or different portions of a target. For instance, in some embodiments, one or more anchor probes may be applied to a sample such that even if not every target is anchored by every anchor probe, a substantial percentage of targets is immobilized to the expandable material. In some cases, for example, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, or at least 90% of the targets are immobilized to the expandable material, e.g., via one or more anchor probes.

Other approaches may be used to anchor nucleic acids besides complementary nucleic acid sequences. For example, in one set of embodiments, the 3'OH end of an RNA molecule may be targeted, e.g., chemically, for immobilization to the expandable material. For example, a 3'OH end may be converted into an aldehyde via oxidation with agents such as periodate (e.g., sodium periodate or potassium periodate). The aldehyde can then be reacted with hydrazines, amines, or other moieties that can render the anchor probe capable of being

incorporated into the expandable material. For example, a hydrazine or an amine can be linked to chemical moieties such as acrylamide, methacrylamide, or other chemicals capable of being incorporated into an expandable material such as polyacrylamide. Additional examples of attachment techniques have been previously described above. Thus, RNA molecules may in some embodiments be anchored only at their 3'OH ends.

As another non-limiting example, the 3'OH could be a site by which enzymatic extension of the RNA could add an anchor probe. For example, polyA polymerase could be used to add modified A nucleotides to the 3' end of the RNA molecule. These A nucleotides can serve as a binding site for polyT probes as described above. As another example, the A nucleotides may be chemically modified such that they contains an appropriate chemical moiety that can be used to anchor the RNA. For example, adenosine nucleotides that contain the click-chemistry reagent azide can be purchased commercially. This could then be reacted with a DBCO group (dibenzocyclooctyne) that is linked to a methacrylate group via an NHS ester, which can then be incorporated into an expandable material, thereby immobilizing the RNA via the 3'OH end to the expandable material.

As still another example, in some embodiments, the 5' end of an RNA molecule may be targeted, e.g., chemically, for immobilization to the expandable material. For example, carbodiimide groups can be used to react with the 5' phosphate present on uncapped RNA molecules. Capped RNA molecules could be decapped using decapping enzymes such as Dcp2 (mRNA-decapping enzyme 2). For example, cross linkers such as EDC (1-ethyl-3-3-dimethyl-aminopropyl carbodiimide) could be reacted with decapped RNAs to specifically label the 5' end of the RNA. These molecules could then be attached to anchor moieties such as acrylamide or methacrylate, e.g., by reaction with methacrylate molecules containing a primary amine. Such moieties can then be incorporated into an expandable material, thereby immobilizing the RNA via the 5' end to the expandable material.

Other molecules, such as proteins, may also be anchored to the expanded matrix, in various embodiments. For example, proteins could be labeled with reagents that contain anchor moieties. A variety of methods can be used for chemical modification of proteins, including crosslinking to amines via aldehydes, crosslinking to cysteines via disulfide bonds or maleimide, etc. Proteins could also be anchored in some embodiments to the expandable gel via interactions with antibodies, which may be labeled in the above fashion.

After immobilization of nucleic acids, or other suitable molecules, to the polymer or gel, other components within the sample may be "cleared." Such clearance may include removal of the components, and/or degradation of the components (e.g., to smaller

components, components that are not fluorescent, etc.) that are not the desired target. In some cases, at least 50%, at least 60%, at least 70%, at least 80%, or at least 90% of the undesired components within the sample may be cleared. Multiple clearance steps can also be performed in certain embodiments, e.g., to remove various undesired components. As discussed, it is believed that the removal of such components may decrease background during analysis (for example, by decreasing background and/or off-target binding), while desired components (such as nucleic acids) can be immobilized and thus not cleared.

For example, proteins may be cleared from the sample using enzymes, denaturants, chelating agents, chemical agents, and the like, which may break down the proteins into smaller components and/or amino acids. These smaller components may be easier to remove physically, and/or may be sufficiently small or inert such that they do not significantly affect the background. Similarly, lipids may be cleared from the sample using surfactants or the like. In some cases, one or more of these are used, e.g., simultaneously or sequentially. Non-limiting examples of suitable enzymes include proteinases such as proteinase K, proteases or peptidases, or digestive enzymes such as trypsin, pepsin, or chymotrypsin. Non-limiting examples of suitable denaturants include guanidine HCl, acetone, acetic acid, urea, or lithium perchlorate. Non-limiting examples of chemical agents able to denature proteins include solvents such as phenol, chloroform, guanidinium isocyananate, urea, formamide, etc. Non-limiting examples of surfactants include Triton X-100 (polyethylene glycol p-(1,1,3,3-tetramethylbutyl)-phenyl ether), SDS (sodium dodecyl sulfate), Igepal CA-630, or poloxamers. Non-limiting examples of chelating agents include ethylenediaminetetraacetic acid (EDTA), citrate, or polyaspartic acid. In some embodiments, compounds such as these may be applied to the sample to clear proteins, lipids, and/or other components. For instance, a buffer solution (e.g., containing Tris or tris(hydroxymethyl)aminomethane) may be applied to the sample, then removed.

Non-limiting examples of DNA enzymes that may be used to remove DNA include DNase I, dsDNase, a variety of restriction enzymes, etc. Non-limiting examples of techniques to clear RNA include RNA enzymes such as RNase A, RNase T, or RNase H, or chemical agents, e.g., via alkaline hydrolysis (for example, by increasing the pH to greater than 10). Non-limiting examples of systems to remove sugars or extracellular matrix include enzymes such as chitinase, heparinases, or other glycosylases. Non-limiting examples of systems to remove lipids include enzymes such as lipidases, chemical agents such as alcohols (e.g., methanol or ethanol), or detergents such as Triton X-100 or sodium dodecyl sulfate. Many of these are readily available commercially. In this way, the background of the sample

may be removed, which may facilitate analysis of the nucleic acid probes or other desired targets, e.g., using fluorescence microscopy, or other techniques as discussed herein. As mentioned, in various embodiments, various targets (e.g., nucleic acids, certain proteins, lipids, viruses, or the like) may be immobilized, while other non-targets may be cleared using suitable agents or enzymes. As a non-limiting example, if a protein (such as an antibody) is immobilized, then RNA enzymes, DNA enzymes, systems to remove lipids, sugars, etc. may be used.

In some cases, the desired target is a nucleic acid. In one set of embodiments, as an illustrative non-limiting example, the sample may be studied by exposing it to one or more types of nucleic acid probes, simultaneously and/or sequentially. For instance, in one set of embodiments, the nucleic acid probes may include smFISH or MERFISH probes, such as those discussed in Int. Pat. Apl. Pub. Nos. WO 2016/018960 or WO 2016/018963, each incorporated herein by reference in its entirety. However, it should be understood that the following is by way of example only, and in other embodiments, the desired target may be, for example, a protein, a lipid, a virus, or the like.

The nucleic acid probes may comprise nucleic acids (or entities that can hybridize to a nucleic acid, e.g., specifically) such as DNA, RNA, LNA (locked nucleic acids), PNA (peptide nucleic acids), or combinations thereof. In some cases, additional components may also be present within the nucleic acid probes, e.g., as discussed below. Any suitable method may be used to introduce nucleic acid probes into a cell or other sample.

For example, in some embodiments, the cell or other sample is fixed prior to introducing the nucleic acid probes, e.g., to preserve the positions of the nucleic acids within the sample. Techniques for fixing cells and tissues are known to those of ordinary skill in the art. As non-limiting examples, a cell may be fixed using chemicals such as formaldehyde, paraformaldehyde, glutaraldehyde, ethanol, methanol, acetone, acetic acid, or the like. In one embodiment, a cell may be fixed using Hepes-glutamic acid buffer-mediated organic solvent (HOPE). The cells may be fixed, in some embodiments, prior to formation and/or after formation of the expandable material. In some cases, the cells are fixed prior to expansion of the expandable material.

The nucleic acid probes may be introduced into the cell (or other sample) using any suitable method. In some cases, the cell may be sufficiently permeabilized such that the nucleic acid probes may be introduced into the cell by flowing a fluid containing the nucleic acid probes around the cells. In some cases, the cells may be sufficiently permeabilized as part of a fixation process; in other embodiments, cells may be permeabilized by exposure to

certain chemicals such as ethanol, methanol, Triton X-100, or the like. In addition, in some embodiments, techniques such as electroporation or microinjection may be used to introduce nucleic acid probes into a cell or other sample.

Certain aspects of the present invention are generally directed to nucleic acid probes that are introduced into a cell (or other sample). The probes may comprise any of a variety of entities that can hybridize to a nucleic acid, typically by Watson-Crick base pairing, such as DNA, RNA, LNA, PNA, etc., depending on the application. The nucleic acid probe typically contains a target sequence that is able to bind to at least a portion of a target nucleic acid, in some cases specifically. When introduced into a cell or other sample, the nucleic acid probe may be able to bind to a specific target nucleic acid (e.g., an mRNA, or other nucleic acids as discussed herein). In some cases, the nucleic acid probes may be determined using signaling entities (e.g., as discussed below), and/or by using secondary nucleic acid probes able to bind to the nucleic acid probes (i.e., to primary nucleic acid probes). The determination of such nucleic acid probes is discussed in detail below.

In some cases, more than one type of (primary) nucleic acid probe may be applied to a sample, e.g., simultaneously. For example, there may be at least 2, at least 5, at least 10, at least 25, at least 50, at least 75, at least 100, at least 300, at least 1,000, at least 3,000, at least 10,000, at least 30,000, at least 50,000, at least 100,000, at least 250,000, at least 500,000, or at least 1,000,000 distinguishable nucleic acid probes that are applied to a sample, e.g., simultaneously or sequentially.

The target sequence may be positioned anywhere within the nucleic acid probe (or primary nucleic acid probe or encoding nucleic acid probe). The target sequence may contain a region that is substantially complementary to a portion of a target nucleic acid. In some cases, the portions may be at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% complementary. In some cases, the target sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 50, at least 60, at least 65, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides in length. In some cases, the target sequence may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than

20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the target sequence may have a length of between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc. Typically, complementarity is determined on the basis of Watson-Crick nucleotide base pairing.

The target sequence of a (primary) nucleic acid probe may be determined with reference to a target nucleic acid suspected of being present within a cell or other sample. For example, a target nucleic acid to a protein may be determined using the protein's sequence, by determining the nucleic acids that are expressed to form the protein. In some cases, only a portion of the nucleic acids encoding the protein are used, e.g., having the lengths as discussed above. In addition, in some cases, more than one target sequence that can be used to identify a particular target may be used. For instance, multiple probes can be used, sequentially and/or simultaneously, that can bind to or hybridize to different regions of the same target. Hybridization typically refers to an annealing process by which complementary single-stranded nucleic acids associate through Watson-Crick nucleotide base pairing (e.g., hydrogen bonding, guanine-cytosine and adenine-thymine) to form double-stranded nucleic acid.

In some embodiments, a nucleic acid probe, such as a primary nucleic acid probe, may also comprise one or more "read" sequences. However, it should be understood that read sequences are not necessary in all cases. In some embodiments, the nucleic acid probe may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 or more, 20 or more, 32 or more, 40 or more, 50 or more, 64 or more, 75 or more, 100 or more, 128 or more read sequences. The read sequences may be positioned anywhere within the nucleic acid probe. If more than one read sequence is present, the read sequences may be positioned next to each other, and/or interspersed with other sequences.

The read sequences, if present, may be of any length. If more than one read sequence is used, the read sequences may independently have the same or different lengths. For instance, the read sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 50, at least 60, at least 65, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides in length. In some cases, the read sequence may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more

than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the read sequence may have a length of between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc.

The read sequence may be arbitrary or random in some embodiments. In certain cases, the read sequences are chosen so as to reduce or minimize homology with other components of the cell or other sample, e.g., such that the read sequences do not themselves bind to or hybridize with other nucleic acids suspected of being within the cell or other sample. In some cases, the homology may be less than 10%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, or less than 1%. In some cases, there may be a homology of less than 20 basepairs, less than 18 basepairs, less than 15 basepairs, less than 14 basepairs, less than 13 basepairs, less than 12 basepairs, less than 11 basepairs, or less than 10 basepairs. In some cases, the basepairs are sequential.

In one set of embodiments, a population of nucleic acid probes may contain a certain number of read sequences, which may be less than the number of targets of the nucleic acid probes in some cases. Those of ordinary skill in the art will be aware that if there is one signaling entity and n read sequences, then in general 2^n-1 different nucleic acid targets may be uniquely identified. However, not all possible combinations need be used. For instance, a population of nucleic acid probes may target 12 different nucleic acid sequences, yet contain no more than 8 read sequences. As another example, a population of nucleic acids may target 140 different nucleic acid species, yet contain no more than 16 read sequences. Different nucleic acid sequence targets may be separately identified by using different combinations of read sequences within each probe. For instance, each probe may contain 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, etc. or more read sequences. In some cases, a population of nucleic acid probes may each contain the same number of read sequences, although in other cases, there may be different numbers of read sequences present on the various probes.

As a non-limiting example, a first nucleic acid probe may contain a first target sequence, a first read sequence, and a second read sequence, while a second, different nucleic acid probe may contain a second target sequence, the same first read sequence, but a third read sequence instead of the second read sequence. Such probes may thereby be distinguished by determining the various read sequences present or associated with a given probe or location, as discussed herein.

In addition, the nucleic acid probes (and their corresponding, complimentary sites on the encoding probes), in certain embodiments, may be made using only 2 or only 3 of the 4 bases, such as leaving out all the "G"s or leaving out all of the "C"s within the probe. Sequences lacking either "G"s or "C"s may form very little secondary structure in certain embodiments, and can contribute to more uniform, faster hybridization.

In some embodiments, the nucleic acid probe may contain a signaling entity. It should be understood that signaling entities are not required in all cases, however; for instance, the nucleic acid probe may be determined using secondary nucleic acid probes in some embodiments, as is discussed in additional detail below. Examples of signaling entities that can be used are also discussed in more detail below.

Other components may also be present within a nucleic acid probe as well. For example, in one set of embodiments, one or more primer sequences may be present, e.g., to allow for enzymatic amplification of probes. Those of ordinary skill in the art will be aware of primer sequences suitable for applications such as amplification (e.g., using PCR or other suitable techniques). Many such primer sequences are available commercially. Other examples of sequences that may be present within a primary nucleic acid probe include, but are not limited to promoter sequences, operons, identification sequences, nonsense sequences, or the like.

Typically, a primer is a single-stranded or partially double-stranded nucleic acid (e.g., DNA) that serves as a starting point for nucleic acid synthesis, allowing polymerase enzymes such as nucleic acid polymerase to extend the primer and replicate the complementary strand. A primer is (e.g., is designed to be) complementary to and to hybridize to a target nucleic acid. In some embodiments, a primer is a synthetic primer. In some embodiments, a primer is a non-naturally-occurring primer. A primer typically has a length of 10 to 50 nucleotides. For example, a primer may have a length of 10 to 40, 10 to 30, 10 to 20, 25 to 50, 15 to 40, 15 to 30, 20 to 50, 20 to 40, or 20 to 30 nucleotides. In some embodiments, a primer has a length of 18 to 24 nucleotides.

In addition, the components of the nucleic acid probe may be arranged in any suitable order. For instance, in one embodiment, the components may be arranged in a nucleic acid probe as: primer—read sequences—targeting sequence—read sequences—reverse primer. The "read sequences" in this structure may each contain any number (including 0) of read sequences, so long as at least one read sequence is present in the probe. Non-limiting example structures include primer—targeting sequence—read sequences—reverse primer, primer—read sequences—targeting sequence—reverse primer, targeting sequence—primer—

targeting sequence—read sequences—reverse primer, targeting sequence—primer—read sequences—targeting sequence—reverse primer, primer—target sequence—read sequences—targeting sequence—reverse primer, targeting sequence—primer—read sequence—reverse primer, targeting sequence—read sequence—primer, read sequence—targeting sequence—primer, read sequence—primer—targeting sequence—reverse primer, etc. In addition, the reverse primer is optional in some embodiments, including in all of the above-described examples.

After introduction of the nucleic acid probes into a cell or other sample, the nucleic acid probes may be directly determined by determining signaling entities (if present), and/or the nucleic acid probes may be determined by using one or more secondary nucleic acid probes, in accordance with certain aspects of the invention. As mentioned, in some cases, the determination may be spatial, e.g., in two or three dimensions. In addition, in some cases, the determination may be quantitative, e.g., the amount or concentration of a primary nucleic acid probe (and of a target nucleic acid) may be determined. Additionally, the secondary probes may comprise any of a variety of entities able to hybridize a nucleic acid, e.g., DNA, RNA, LNA, and/or PNA, etc., depending on the application. Signaling entities are discussed in more detail below.

A secondary nucleic acid probe may contain a recognition sequence able to bind to or hybridize with a read sequence of a primary nucleic acid probe. In some cases, the binding is specific, or the binding may be such that a recognition sequence preferentially binds to or hybridizes with only one of the read sequences that are present. The secondary nucleic acid probe may also contain one or more signaling entities. If more than one secondary nucleic acid probe is used, the signaling entities may be the same or different.

The recognition sequences may be of any length, and multiple recognition sequences may be of the same or different lengths. If more than one recognition sequence is used, the recognition sequences may independently have the same or different lengths. For instance, the recognition sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, or at least 50 nucleotides in length. In some cases, the recognition sequence may be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the recognition sequence may have a length of between 10 and 30, between 20 and 40, or between 25 and 35 nucleotides, etc. In one embodiment, the recognition sequence is of the same length as the read sequence. In

addition, in some cases, the recognition sequence may be at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 100% complementary to a read sequence of the primary nucleic acid probe.

As mentioned, in some cases, the secondary nucleic acid probe may comprise one or more signaling entities. Examples of signaling entities are discussed in more detail below.

As discussed, in certain aspects of the invention, nucleic acid probes are used that contain various "read sequences." For example, a population of primary nucleic acid probes may contain certain "read sequences" which can bind certain of the secondary nucleic acid probes, and the locations of the primary nucleic acid probes are determined within the sample using secondary nucleic acid probes, e.g., which comprise a signaling entity. As mentioned, in some cases, a population of read sequences may be combined in various combinations to produce different nucleic acid probes, e.g., such that a relatively small number of read sequences may be used to produce a relatively large number of different nucleic acid probes.

Thus, in some cases, a population of primary nucleic acid probes (or other nucleic acid probes) may each contain a certain number of read sequences, some of which are shared between different primary nucleic acid probes such that the total population of primary nucleic acid probes may contain a certain number of read sequences. A population of nucleic acid probes may have any suitable number of read sequences. For example, a population of primary nucleic acid probes may have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 etc. read sequences. More than 20 are also possible in some embodiments. In addition, in some cases, a population of nucleic acid probes may, in total, have 1 or more, 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, 12 or more, 13 or more, 14 or more, 15 or more, 16 or more, 20 or more, 24 or more, 32 or more, 40 or more, 50 or more, 60 or more, 64 or more, 100 or more, 128 or more, etc. of possible read sequences present, although some or all of the probes may each contain more than one read sequence, as discussed herein. In addition, in some embodiments, the population of nucleic acid probes may have no more than 100, no more than 80, no more than 64, no more than 60, no more than 50, no more than 40, no more than 32, no more than 24, no more than 20, no more than 16, no more than 15, no more than 14, no more than 13, no more than 12, no more than 11, no more than 10, no more than 9, no more than 8, no more than 7, no more than 6, no more than 5, no more than 4, no more than 3, or no more than two read sequences present. Combinations of any of these are also possible, e.g., a population of nucleic acid probes may comprise between 10 and 15 read sequences in total.

As a non-limiting example of an approach to combinatorially producing a relatively large number of nucleic acid probes from a relatively small number of read sequences, in a population of 6 different types of nucleic acid probes, each comprising one or more read sequences, the total number of read sequences within the population may be no greater than 4. It should be understood that although 4 read sequences are used in this example for ease of explanation, in other embodiments, larger numbers of nucleic acid probes may be realized, for example, using 5, 8, 10, 16, 32, etc. or more read sequences, or any other suitable number of read sequences described herein, depending on the application. If each of the primary nucleic acid probes contains two different read sequences, then by using 4 such read sequences (A, B, C, and D), up to 6 probes may be separately identified. It should be noted that in this example, the ordering of read sequences on a nucleic acid probe is not essential, i.e., "AB" and "BA" may be treated as being synonymous (although in other embodiments, the ordering of read sequences may be essential and "AB" and "BA" may not necessarily be synonymous). Similarly, if 5 read sequences are used (A, B, C, D, and E) in the population of primary nucleic acid probes, up to 10 probes may be separately identified. For example, one of ordinary skill in the art would understand that, for k read sequences in a population with n read sequences on each probe, up to $\binom{n}{k}$ different probes may be produced, assuming that the ordering of read sequences is not essential; because not all of the probes need to have the same number of read sequences and not all combinations of read sequences need to be used in every embodiment, either more or less than this number of different probes may also be used in certain embodiments. In addition, it should also be understood that the number of read sequences on each probe need not be identical in some embodiments. For instance example, some probes may contain 2 read sequences while other probes may contain 3 read sequences.

In some aspects, the read sequences and/or the pattern of binding of nucleic acid probes within a sample may be used to define an error-detecting and/or an error-correcting code, for example, to reduce or prevent misidentification or errors of the nucleic acids. Thus, for example, if binding is indicated (e.g., as determined using a signaling entity), then the location may be identified with a "1"; conversely, if no binding is indicated, then the location may be identified with a "0" (or vice versa, in some cases). Multiple rounds of binding determinations, e.g., using different nucleic acid probes, can then be used to create a "codeword," e.g., for that spatial location. In some embodiments, the codeword may be subjected to error detection and/or correction. For instance, the codewords may be organized

such that, if no match is found for a given set of read sequences or binding pattern of nucleic acid probes, then the match may be identified as an error, and optionally, error correction may be applied sequences to determine the correct target for the nucleic acid probes. In some cases, the codewords may have fewer "letters" or positions that the total number of nucleic acids encoded by the codewords, e.g. where each codeword encodes a different nucleic acid.

Such error-detecting and/or the error-correction code may take a variety of forms. A variety of such codes have previously been developed in other contexts such as the telecommunications industry, such as Golay codes or Hamming codes. In one set of embodiments, the read sequences or binding patterns of the nucleic acid probes are assigned such that not every possible combination is assigned.

For example, if 4 read sequences are possible and a primary nucleic acid probe contains 2 read sequences, then up to 6 primary nucleic acid probes could be identified; but the number of primary nucleic acid probes used may be less than 6. Similarly, for k read sequences in a population with n read sequences on each primary nucleic acid probe, $\binom{n}{k}$ different probes may be produced, but the number of primary nucleic acid probes that are used may be any number more or less than $\binom{n}{k}$. In addition, these may be randomly assigned, or assigned in specific ways to increase the ability to detect and/or correct errors.

As another example, if multiple rounds of nucleic acid probes are used, the number of rounds may be arbitrarily chosen. If in each round, each target can give two possible outcomes, such as being detected or not being detected, up to 2^n different targets may be possible for n rounds of probes, but the number of nucleic acid targets that are actually used may be any number less than 2^n. For example, if in each round, each target can give more than two possible outcomes, such as being detected in different color channels, more than 2^n (e.g. 3^n, 4^n ...) different targets may be possible for n rounds of probes. In some cases, the number of nucleic acid targets that are actually used may be any number less than this number. In addition, these may be randomly assigned, or assigned in specific ways to increase the ability to detect and/or correct errors.

For example, in one set of embodiments, the codewords or nucleic acid probes may be assigned within a code space such that the assignments are separated by a Hamming distance, which measures the number of incorrect "reads" in a given pattern that cause the nucleic acid probe to be misinterpreted as a different valid nucleic acid probe. In certain cases, the Hamming distance may be at least 2, at least 3, at least 4, at least 5, at least 6, or the like. In addition, in one set of embodiments, the assignments may be formed as a Hamming code, for

instance, a Hamming(7, 4) code, a Hamming(15, 11) code, a Hamming(31, 26) code, a Hamming(63, 57) code, a Hamming(127, 120) code, etc. In another set of embodiments, the assignments may form a SECDED code, e.g., a SECDED(8,4) code, a SECDED(16,4) code, a SCEDED(16, 11) code, a SCEDED(22, 16) code, a SCEDED(39, 32) code, a SCEDED(72, 64) code, etc. In yet another set of embodiments, the assignments may form an extended binary Golay code, a perfect binary Golay code, or a ternary Golay code. In another set of embodiments, the assignments may represent a subset of the possible values taken from any of the codes described above.

For example, a code with the same error correcting properties of the SECDED code may be formed by using only binary words that contain a fixed number of '1' bits, such as 4, to encode the targets. In another set of embodiments, the assignments may represent a subset of the possible values taken from codes described above for the purpose of addressing asymmetric readout errors. For example, in some cases, a code in which the number of '1' bits may be fixed for all used binary words may eliminate the biased measurement of words with different numbers of '1's when the rate at which '0' bits are measured as '1's or '1' bits are measured as '0's are different.

Accordingly, in some embodiments, once the codeword is determined (e.g., as discussed herein), the codeword may be compared to the known nucleic acid codewords. If a match is found, then the nucleic acid target can be identified or determined. If no match is found, then an error in the reading of the codeword may be identified. In some cases, error correction can also be applied to determine the correct codeword, and thus resulting in the correct identity of the nucleic acid target. In some cases, the codewords may be selected such that, assuming that there is only one error present, only one possible correct codeword is available, and thus, only one correct identity of the nucleic acid target is possible. In some cases, this may also be generalized to larger codeword spacings or Hamming distances; for instance, the codewords may be selected such that if two, three, or four errors are present (or more in some cases), only one possible correct codeword is available, and thus, only one correct identity of the nucleic acid targets is possible.

The error-correcting code may be a binary error-correcting code, or it may be based on other numbering systems, e.g., ternary or quaternary error-correcting codes. For instance, in one set of embodiments, more than one type of signaling entity may be used and assigned to different numbers within the error-correcting code. Thus, as a non-limiting example, a first signaling entity (or more than one signaling entity, in some cases) may be assigned as "1" and a second signaling entity (or more than one signaling entity, in some cases) may be

assigned as "2" (with "0" indicating no signaling entity present), and the codewords distributed to define a ternary error-correcting code. Similarly, a third signaling entity may additionally be assigned as "3" to make a quaternary error-correcting code, etc.

As discussed above, in certain aspects, signaling entities are determined, e.g., to determine nucleic acid probes and/or to create codewords. In some cases, signaling entities within a sample may be determined, e.g., spatially, using a variety of techniques. In some embodiments, the signaling entities may be fluorescent, and techniques for determining fluorescence within a sample, such as fluorescence microscopy or confocal microscopy, may be used to spatially identify the positions of signaling entities within a cell. In some cases, the positions of entities within the sample may be determined in two or even three dimensions. In addition, in some embodiments, more than one signaling entity may be determined at a time (e.g., signaling entities with different colors or emissions), and/or sequentially.

In addition, in some embodiments, a confidence level for the identified nucleic acid target may be determined. For example, the confidence level may be determined using a ratio of the number of exact matches to the number of matches having one or more one-bit errors. In some cases, only matches having a confidence ratio greater than a certain value may be used. For instance, in certain embodiments, matches may be accepted only if the confidence ratio for the match is greater than about 0.01, greater than about 0.03, greater than about 0.05, greater than about 0.1, greater than about 0.3, greater than about 0.5, greater than about 1, greater than about 3, greater than about 5, greater than about 10, greater than about 30, greater than about 50, greater than about 100, greater than about 300, greater than about 500, greater than about 1000, or any other suitable value. In addition, in some embodiments, matches may be accepted only if the confidence ratio for the identified nucleic acid target is greater than an internal standard or false positive control by about 0.01, about 0.03, about 0.05, about 0.1, about 0.3, about 0.5, about 1, about 3, about 5, about 10, about 30, about 50, about 100, about 300, about 500, about 1000, or any other suitable value

In some embodiments, the spatial positions of the entities (and thus, nucleic acid probes that the entities may be associated with) may be determined at relatively high resolutions. For instance, the positions may be determined at spatial resolutions of better than about 100 micrometers, better than about 30 micrometers, better than about 10 micrometers, better than about 3 micrometers, better than about 1 micrometer, better than about 800 nm, better than about 600 nm, better than about 500 nm, better than about 400 nm, better than about 300 nm, better than about 200 nm, better than about 100 nm, better than about 90 nm,

better than about 80 nm, better than about 70 nm, better than about 60 nm, better than about 50 nm, better than about 40 nm, better than about 30 nm, better than about 20 nm, or better than about 10 nm, etc.

There are a variety of techniques able to determine or image the spatial positions of entities or targets optically, e.g., using fluorescence microscopy, using radioactivity, via conjugation with suitable chromophores, or the like. For example, various conventional microscopy techniques that may be used in various embodiments of the invention include, but are not limited to, epi-fluorescence microscopy, total-internal-reflectance microscopy, highly-inclined thin-illumination (HILO) microscopy, light-sheet microscopy, scanning confocal microscopy, scanning line confocal microscopy, spinning disk confocal microscopy, or other comparable conventional microscopy techniques.

In some embodiments, *in situ* hybridization (ISH) techniques for labeling nucleic acids such as DNA or RNA may be used, e.g., where nucleic acid probes may be hybridized to nucleic acids in samples. These may be performed, e.g., at cellular-scale or single-molecule-scale resolutions. In some cases, the ISH probes can be composed of RNA, DNA, PNA, LNA, other synthetic nucleotides, or the like, and/or a combination of any of these. The presence of a hybridized probe can be measured, for example, with radioactivity using radioactively labeled nucleic acid probes, immunohistochemistry using, for example, biotin labeled nucleic acid probes, enzymatic chromophore or fluorophore generation using, for example, probes that can bind enzymes such as horseradish peroxidase and approaches such as tyramide signal amplification, fluorescence imaging using nucleic acid probes directly labeled with fluorophores, or hybridization of secondary nucleic acid probes to these primary probes, with the secondary probes detected via any of the above methods.

In some cases, the spatial positions may be determined at super resolutions, or at resolutions better than the wavelength of light or the diffraction limit (although in other embodiments, super resolutions are not required). Non-limiting examples include STORM (stochastic optical reconstruction microscopy), STED (stimulated emission depletion microscopy), NSOM (Near-field Scanning Optical Microscopy), 4Pi microscopy, SIM (Structured Illumination Microscopy), SMI (Spatially Modulated Illumination) microscopy, RESOLFT (Reversible Saturable Optically Linear Fluorescence Transition Microscopy), GSD (Ground State Depletion Microscopy), SSIM (Saturated Structured-Illumination Microscopy), SPDM (Spectral Precision Distance Microscopy), Photo-Activated Localization Microscopy (PALM), Fluorescence Photoactivation Localization Microscopy (FPALM), LIMON (3D Light Microscopical Nanosizing Microscopy), Super-resolution optical

fluctuation imaging (SOFI), or the like. See, e.g., U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*; or Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*, each incorporated herein by reference in their entireties.

In one embodiment, the sample may be illuminated by single Gaussian mode laser lines. In some embodiments, the illumination profiled may be flattened by passing these laser lines through a multimode fiber that is vibrated via piezo-electric or other mechanical means. In some embodiments, the illumination profile may be flattened by passing single-mode, Gaussian beams through a variety of refractive beam shapers, such as the piShaper or a series of stacked Powell lenses. In yet another set of embodiments, the Gaussian beams may be passed through a variety of different diffusing elements, such as ground glass or engineered diffusers, which may be spun in some cases at high speeds to remove residual laser speckle. In yet another embodiment, laser illumination may be passed through a series of lenslet arrays to produce overlapping images of the illumination that approximate a flat illumination field.

In some embodiments, the centroids of the spatial positions of the entities may be determined. For example, a centroid of a signaling entity may be determined within an image or series of images using image analysis algorithms known to those of ordinary skill in the art. In some cases, the algorithms may be selected to determine non-overlapping single emitters and/or partially overlapping single emitters in a sample. Non-limiting examples of suitable techniques include a maximum likelihood algorithm, a least squares algorithm, a Bayesian algorithm, a compressed sensing algorithm, or the like. Combinations of these techniques may also be used in some cases.

In addition, the signaling entity may be inactivated in some cases. For example, in some embodiments, a first secondary nucleic acid probe containing a signaling entity may be applied to a sample that can recognize a first read sequence, then the first secondary nucleic acid probe can be inactivated before a second secondary nucleic acid probe is applied to the sample. If multiple signaling entities are used, the same or different techniques may be used to inactivate the signaling entities, and some or all of the multiple signaling entities may be inactivated, e.g., sequentially or simultaneously.

Inactivation may be caused by removal of the signaling entity (e.g., from the sample, or from the nucleic acid probe, etc.), and/or by chemically altering the signaling entity in

some fashion, e.g., by photobleaching the signaling entity, bleaching or chemically altering the structure of the signaling entity, e.g., by reduction, etc.). For instance, in one set of embodiments, a fluorescent signaling entity may be inactivated by chemical or optical techniques such as oxidation, photobleaching, chemically bleaching, stringent washing or enzymatic digestion or reaction by exposure to an enzyme, dissociating the signaling entity from other components (e.g., a probe), chemical reaction of the signaling entity (e.g., to a reactant able to alter the structure of the signaling entity) or the like. For instance, bleaching may occur by exposure to oxygen, reducing agents, or the signaling entity could be chemically cleaved from the nucleic acid probe and washed away via fluid flow.

In some embodiments, various nucleic acid probes (including primary and/or secondary nucleic acid probes) may include one or more signaling entities. If more than one nucleic acid probe is used, the signaling entities may each by the same or different. In certain embodiments, a signaling entity is any entity able to emit light. For instance, in one embodiment, the signaling entity is fluorescent. In other embodiments, the signaling entity may be phosphorescent, radioactive, absorptive, etc. In some cases, the signaling entity is any entity that can be determined within a sample at relatively high resolutions, e.g., at resolutions better than the wavelength of visible light or the diffraction limit. The signaling entity may be, for example, a dye, a small molecule, a peptide or protein, or the like. The signaling entity may be a single molecule in some cases. If multiple secondary nucleic acid probes are used, the nucleic acid probes may comprise the same or different signaling entities.

Non-limiting examples of signaling entities include fluorescent entities (fluorophores) or phosphorescent entities, for example, cyanine dyes (e.g., Cy2, Cy3, Cy3B, Cy5, Cy5.5, Cy7, etc.), Alexa Fluor dyes, Atto dyes, photoswtichable dyes, photoactivatable dyes, fluorescent dyes, metal nanoparticles, semiconductor nanoparticles or "quantum dots", fluorescent proteins such as GFP (Green Fluorescent Protein), or photoactivabale fluorescent proteins, such as PAGFP, PSCFP, PSCFP2, Dendra, Dendra2, EosFP, tdEos, mEos2, mEos3, PAmCherry, PAtagRFP, mMaple, mMaple2, and mMaple3. Other suitable signaling entities are known to those of ordinary skill in the art. See, e.g., U.S. Pat. No. 7,838,302 or U.S. Pat. Apl. Ser. No. 61/979,436, each incorporated herein by reference in its entirety. In some cases, spectrally distinct fluorescent dyes may be used.

In one set of embodiments, the signaling entity may be attached to an oligonucleotide sequence via a bond that can be cleaved to release the signaling entity. In one set of embodiments, a fluorophore may be conjugated to an oligonucleotide via a cleavable bond,

such as a photocleavable bond. Non-limiting examples of photocleavable bonds include, but are not limited to, 1-(2-nitrophenyl)ethyl, 2□nitrobenzyl, biotin phosphoramidite, acrylic phosphoramidite, diethylaminocoumarin, 1-(4,5-dimethoxy-2-nitrophenyl)ethyl, cyclo-dodecyl (dimethoxy-2-nitrophenyl)ethyl, 4-aminomethyl-3-nitrobenzyl, (4-nitro-3-(1-chlorocarbonyloxyethyl)phenyl)methyl-S-acetylthioic acid ester, (4-nitro-3-(1-thlorocarbonyloxyethyl)phenyl)methyl-3-(2-pyridyldithiopropionic acid) ester, 3-(4,4'-dimethoxytrityl)-1-(2-nitrophenyl)-propane-1,3-diol-[2- cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-trifluoroacetylcaproamidomethyl)phenyl]-ethyl-[2-cyano-ethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-(4,4'-dimethoxytrityloxy)butyramidomethyl)phenyl]-ethyl-[2-cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-(N-(4,4'-dimethoxytrityl))-biotinamidocaproamido-methyl)phenyl]-ethyl-[2-cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, or similar linkers. In another set of embodiments, the fluorophore may be conjugated to an oligonucleotide via a disulfide bond. The disulfide bond may be cleaved by a variety of reducing agents such as, but not limited to, dithiothreitol, dithioerythritol, beta-mercaptoethanol, sodium borohydride, thioredoxin, glutaredoxin, trypsinogen, hydrazine, diisobutylaluminum hydride, oxalic acid, formic acid, ascorbic acid, phosphorous acid, tin chloride, glutathione, thioglycolate, 2,3-dimercaptopropanol, 2-mercaptoethylamine, 2-aminoethanol, tris(2-carboxyethyl)phosphine, bis(2-mercaptoethyl) sulfone, N,N'-dimethyl-N,N'-bis(mercaptoacetyl)hydrazine, 3-mercaptoproptionate, dimethylformamide, thiopropyl-agarose, tri-*n*-butylphosphine, cysteine, iron sulfate, sodium sulfite, phosphite, hypophosphite, phosphorothioate, or the like, and/or combinations of any of these. In another embodiment, the fluorophore may be conjugated to an oligonucleotide via one or more phosphorothioate modified nucleotides in which the sulfur modification replaces the bridging and/or non-bridging oxygen. The fluorophore may be cleaved from the oligonucleotide, in certain embodiments, via addition of compounds such as but not limited to iodoethanol, iodine mixed in ethanol, silver nitrate, or mercury chloride. In yet another set of embodiments, the signaling entity may be chemically inactivated through reduction or oxidation. For example, in one embodiment, a chromophore such as Cy5 or Cy7 may be reduced using sodium borohydride to a stable, non-fluorescence state. In still another set of embodiments, a fluorophore may be conjugated to an oligonucleotide via an azo bond, and the azo bond may be cleaved with 2-[(2-N-arylamino)phenylazo]pyridine. In yet another set of embodiments, a fluorophore may be conjugated to an oligonucleotide via a suitable nucleic acid segment that can be cleaved upon suitable exposure to DNAse, e.g., an exodeoxyribonuclease or an endodeoxyribonuclease. Examples include, but are not limited

to, deoxyribonuclease I or deoxyribonuclease II. In one set of embodiments, the cleavage may occur via a restriction endonuclease. Non-limiting examples of potentially suitable restriction endonucleases include BamHI, BsrI, NotI, XmaI, PspAI, DpnI, MboI, MnlI, Eco57I, Ksp632I, DraIII, AhaII, SmaI, MluI, HpaI, ApaI, BclI, BstEII, TaqI, EcoRI, SacI, HindII, HaeII, DraII, Tsp509I, Sau3AI, PacI, etc. Over 3000 restriction enzymes have been studied in detail, and more than 600 of these are available commercially. In yet another set of embodiments, a fluorophore may be conjugated to biotin, and the oligonucleotide conjugated to avidin or streptavidin. An interaction between biotin and avidin or streptavidin allows the fluorophore to be conjugated to the oligonucleotide, while sufficient exposure to an excess of addition, free biotin could "outcompete" the linkage and thereby cause cleavage to occur. In addition, in another set of embodiments, the probes may be removed using corresponding "toe-hold-probes," which comprise the same sequence as the probe, as well as an extra number of bases of homology to the encoding probes (e.g., 1-20 extra bases, for example, 5 extra bases). These probes may remove the labeled readout probe through a strand-displacement interaction.

As used herein, the term "light" generally refers to electromagnetic radiation, having any suitable wavelength (or equivalently, frequency). For instance, in some embodiments, the light may include wavelengths in the optical or visual range (for example, having a wavelength of between about 400 nm and about 700 nm, i.e., "visible light"), infrared wavelengths (for example, having a wavelength of between about 300 micrometers and 700 nm), ultraviolet wavelengths (for example, having a wavelength of between about 400 nm and about 10 nm), or the like. In certain cases, as discussed in detail below, more than one entity may be used, i.e., entities that are chemically different or distinct, for example, structurally. However, in other cases, the entities may be chemically identical or at least substantially chemically identical.

Another aspect of the invention is directed to a computer-implemented method. For instance, a computer and/or an automated system may be provided that is able to automatically and/or repetitively perform any of the methods described herein. As used herein, "automated" devices refer to devices that are able to operate without human direction, i.e., an automated device can perform a function during a period of time after any human has finished taking any action to promote the function, e.g. by entering instructions into a computer to start the process. Typically, automated equipment can perform repetitive functions after this point in time. The processing steps may also be recorded onto a machine-readable medium in some cases.

For example, in some cases, a computer may be used to control imaging of the sample, e.g., using fluorescence microscopy, STORM or other super-resolution techniques such as those described herein. In some cases, the computer may also control operations such as drift correction, physical registration, hybridization and cluster alignment in image analysis, cluster decoding (e.g., fluorescent cluster decoding), error detection or correction (e.g., as discussed herein), noise reduction, identification of foreground features from background features (such as noise or debris in images), or the like. As an example, the computer may be used to control activation and/or excitation of signaling entities within the sample, and/or the acquisition of images of the signaling entities. In one set of embodiments, a sample may be excited using light having various wavelengths and/or intensities, and the sequence of the wavelengths of light used to excite the sample may be correlated, using a computer, to the images acquired of the sample containing the signaling entities. For instance, the computer may apply light having various wavelengths and/or intensities to a sample to yield different average numbers of signaling entities in each region of interest (e.g., one activated entity per location, two activated entities per location, etc.). In some cases, this information may be used to construct an image and/or determine the locations of the signaling entities, in some cases at high resolutions, as noted above.

The following documents are incorporated herein by reference in their entireties: International Patent Application No. PCT/US2007/017618, filed August 7, 2007, entitled "Sub-Diffraction Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*, published as WO 2008/091296 on July 31, 2008; U.S. Patent No. 7,776,613, issued August 17, 2010, entitled "Sub-Diffraction Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Patent No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; International Patent Application No. PCT/US2015/042559, filed July 29, 2015, entitled "Probe Library Construction," by Zhuang, *et al.*, published as WO 2016/018963 on February 4, 2016; International Patent Application No. PCT/US2015/042556, filed July 29, 2015, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*, published as WO 2016/018960 on February 4, 2016; U.S. Patent Application Serial No. 15/329,683, filed January 27, 2017, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*, published as U.S. Patent Application Publication No. 2017/0220733 on August 3, 2017; U.S. Patent Application Serial No. 15/329,651, filed January 27, 2017, entitled "Probe Library Construction," by Zhuang, *et al.*, published as U.S. Patent Application Publication No. 2017/0212986 on July 27, 2017; U.S. Patent Application Serial

No. 15/252,307, filed August 31, 2016, entitled "Sub-Diffraction Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*, published as U.S. Patent Application Publication No. 2016/0370295 on December 22, 2016; U.S. Provisional Patent Application Serial No. 62/419,033, filed November 8, 2016, entitled "Matrix Imprinting and Clearing," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/511,920, filed May 26, 2017, entitled "Systems and Methods for High-Throughput Image-Based Screening," by Zhuang, *et al.*; and U.S. Provisional Patent Application Serial No. 62/569,127, filed October 6, 2017, entitled "Multiplexed Imaging Using MERFISH, Expansion Microscopy, and Related Technologies," by Zhuang, *et al.*

The following examples are intended to illustrate certain embodiments of the present invention, but do not exemplify the full scope of the invention.

EXAMPLE 1

Some of the following examples illustrates expansion MERFISH, based on certain embodiments of the invention, which exploit expansion microscopy to substantially increase the total density of RNAs measurable by MERFISH. In these examples, RNAs are anchored to a polymer gel that were expanded 12-fold in volume to physically lower RNA densities. With this, accurate identification of RNAs was demonstrated in a MERFISH library composed of abundant RNAs, the total density of which was >10 fold higher than previously measured RNA libraries.

In addition, some of the following examples illustrates combining immunofluorescence with expansion MERFISH, based on certain embodiments, which exploit acrydite-modified oligo-conjugated antibodies to reveal protein localization and expression together with measurements of RNAs. In these examples, samples were stained with acrydite-modified oligonucleotides-conjugated antibodies prior to proteolysis, providing crosslinking to the expandable gel, and then digested. The linked oligonucleotides were stained to reveal the original location of the antibodies together with MERFISH readout of RNAs.

In situ imaging-based approaches to single-cell transcriptomics allow not only the expression profile of individual cells to be determined, but also the spatial positions of individual RNA molecules to be localized. These approaches provide powerful means to map the spatial organizations of RNAs inside cells and the transcriptionally distinct cells in tissues. Currently available image-based single-cell RNA profiling methods rely on either multiplexed fluorescence *in situ* hybridization (FISH) or in situ sequencing. In particular, multiplexed error-robust FISH (MERFISH), a massively multiplexed form of single-molecule

FISH (smFISH), allows RNA imaging at the transcriptomic scale. As a powerful method that images individual RNA molecules inside cells, smFISH provides the precise copy number and spatial organization of RNAs in single cells. MERFISH multiplexes smFISH measurements by labeling RNAs combinatorically with oligonucleotide probes which contain error-robust barcodes and measuring these barcodes through sequential rounds of smFISH imaging. Using this approach, simultaneous imaging of hundreds to thousands of RNA species in individual cells using error detection/correction barcoding schemes have been demonstrated. See, e.g., U.S. Pat. Apl. Pub. No. 2017-0220733, entitled "Systems and Methods for Determining Nucleic Acids," and U.S. Pat. Apl. Pub. No. 2017-0212986, entitled "Probe Library Construction," each incorporated herein by reference in its entirety.

The measurement throughput of MERFISH to tens of thousands of cells per single-day-long measurement may also be increased. In addition, sample clearing approaches that increase the signal-to-background ratio by anchoring cellular RNAs to a polymer matrix and removing other cellular components that give rise to fluorescence background have been developed, and this clearing approaches allows high-quality MERFISH measurement of tissue sections. See, e.g., U.S. Pat. Apl. Pub. No. 62/419,033, entitled "Matrix Imprinting and Clearing"; and U.S. Pat. Apl. Pub. No. 62/511,920, entitled "Systems and methods for high-throughput image-based screening," each incorporated herein by reference in its entirety.

In order to accurately identify RNA molecules, MERFISH, as well as other multiplexed FISH or in situ sequencing based RNA profiling methods, requires non-overlapping fluorescence signals from individual RNAs. However, due to the diffraction limit, when molecules are sufficiently close to each other, their fluorescent signals will overlap, limiting the density of RNAs that can be imaged simultaneously. This problem can be overcome, for example, by super-resolution imaging methods, either through optical means or through sample expansion. These examples illustrates sample expansion and expansion microscopy (ExM) to substantially improve the ability to resolve nearby molecules.

In this approach, the desired signal is conjugated to an expandable polyacrylamide gel, and then the gel is physically expanded by changing the ionic strength of the buffer, separating molecules that would have otherwise produced overlapping fluorescent signals. These examples demonstrates an approach to combine MERFISH and expansion microscopy. The mRNAs is anchored to an expandable polymer gel though acrydite-modified poly-dT oligonucleotides and imaged a high-abundance RNA library in cultured human osteosarcoma

cells (U-2 OS), which contains ~130 RNA species. Without gel expansion, these RNAs were not well resolved and hence detected with a relatively low detection efficiency of ~15-20%. In contrast, in expanded sample, individual RNA molecules became well resolved, leading to a substantial increase in their detection efficiency. Comparison with smFISH and bulk sequencing results demonstrated that these RNAs in the expanded sample were detected with high accuracy and near 100% efficiency. These examples also demonstrated the ability to do simultaneous MERFISH RNA imaging and immunofluorescence imaging of proteins in these expanded samples.

EXAMPLE 2

Effect of RNA density on the detection efficiency of MERFISH measurements. To illustrate the effect of RNA density on multiplexed smFISH measurements, a high-abundance 129-RNA library was measured using a previously published 16-bit modified Hamming distance 4 (MHD4) binary codes, which allows error detection and correction. This code includes 140 unique code words, and ~129 of them were used to encode the RNAs. ~11 were used as blank controls that do not correspond to any RNA. Among the 129 targeted RNAs, 106 were in the abundance range of 40 – 250 copies per cell, and the remaining 23 spanned an abundance range of 1 – 1000 copies per cell to quantify performance across different abundances. The total abundance of RNAs in this library was 14-fold higher than a 130-RNA library previously measured using the MHD4 code with ~80-90% detection efficiency.

In the MERFISH measurements, the RNAs were labeled with two sets of probes. In the first step, each cellular RNA was hybridized with a complex set of oligonucleotide probes termed "encoding probes," which contained targeting sequences that bind cellular RNAs and readout sequences that determines the barcodes of these RNAs. In second step, the readout sequences, and hence the barcodes, were detected through a series of smFISH measurements, each round with one or more readout probe complementary to one or more readout sequence. MERFISH measurements were carried out in U-2 OS cells using matrix-imprinting-based clearing methods. See U.S. Pat. Apl. Pub. No. 62/419,033, entitled "Matrix Imprinting and Clearing," incorporated herein by reference. Briefly, the cells were fixed, permeabilized, labeled with encoding probes to the 129 RNA species as well as acrydite-modified poly-dT probes that target polyadenylated (polyA) RNAs. The cells were embedded in a polymer gel and cellular proteins and lipids removed by Proteinase K digestion and detergent extraction, while the polyA RNAs were anchored to the gel through the poly-dT probes. After the clearing, eight rounds of two-color smFISH measurements, each with two readout probes, were carried out to read out the 16-bit barcodes on the RNAs, chemical cleavage was used to

remove the fluorophores that were linked to the readout probes between consecutive rounds of smFISH imaging. Because the height of the cells were greater than the thickness of a single optical section, the sample was imaged with multiple z-sections to ensure that >90% RNA molecules within the cells were detected.

Unlike previous MERFISH measurements on lower abundance RNA libraries because of the high molecular density associated with this 14-fold higher abundance RNA library, a substantial fraction of smFISH signals overlap in space in each round of imaging (Fig. 1A, B). As a result, only a small fraction of the RNA molecules were decodable (Fig. 1C). Comparison with bulk RNA-seq shows that the average copy number per cell detected for these RNAs by MERFISH correlated with the RNA abundance measured by RNA-seq with a Pearson correlation coefficient of 0.6 between the $\log_{10}$ values (Fig. 1D). To determine the detection efficiency, the MERFISH results were compared with the smFISH measurements for 12 genes in this library, also carried out using the matrix-imprinting-based clearing method. This comparison showed that the copy numbers per cell determined by MERFISH were only 21% +/- 4% (average +/- SEM) or 16% (median) of those determined by smFISH (Fig. 1E).

Fig. 1 shows MERFISH measurements of a high-abundance RNA library in unexpanded U-2 OS samples. The samples were cleared using the matrix-imprinting-based clearing method as described above. Fig. 1A shows an image of MERFISH measurement in an unexpanded U-2 OS sample stained with encoding probes for 129 RNAs and visualized with a Cy5-labeled readout probe at one focal plane of the z-scan. The solid lines mark the edge of a cell and the dashed lines mark the DAPI-stained region of a nucleus. Fig. 1B shows high-pass filtered fluorescence images of all 8 rounds of two-color smFISH imaging for the boxed sub-region in (Fig. 1A). Different shadings represent the Cy5 channel (green) and Alexa 750 (red) channel, respectively. Fig. 1C shows the localizations of all decoded RNAs in Fig. 1A based on their measured binary barcodes. The inset shows the localizations of all decoded RNAs of the region shown in Fig. 1B. Decoded RNAs across all optical sections are displayed. The solid lines mark the edge of a cell and the dashed lines mark the DAPI-stained region of a nucleus. Fig. 1D shows the average RNA copy numbers per cell for the 129 RNA species determined by MERFISH vs. the abundances as determined by RNA-seq. The Pearson correlation coefficient between the $\log_{10}$ values (ρ_{10}) is 0.6. Fig. 1E shows the average RNA copy numbers per cell determined by MERFISH vs. those by smFISH measurements for 12 of the 129 RNAs. A z-scan of 5 micrometers in depth was performed for each dataset to ensure that at least 90% of the RNA molecules were included in the

imaged depth. The Pearson correlation coefficient between the $\log_{10}$ values (ρ_{10}) is 0.75. The mean ratio of the copy number values determined by MERFISH to that determined by smFISH is 21% +/- 4% (SEM, n = 12 RNA species) and the median 16%.

EXAMPLE 3

High RNA-density MERFISH measurements with expansion microscopy. It is believed that the reduced MERFISH detection efficiency may be due, in part, to overlapping single-molecule signals. This problem can be overcome, as is shown in this example, by super-resolution imaging methods, either through optical means or through sample expansion. This example illustrates sample expansion with molecules anchored at a single location and expansion microscopy (ExM) to increase the distance between molecules and substantially improve the ability to resolve nearby molecules.

In this example, cells were fixed and permeabilized, and labeled with encoding probes and poly-dT anchoring probes as described above. Then the cells were embedded in an expandable polymer gel. Afterwards, cells were digested by Proteinase K to remove proteins and homogenize the mechanical properties of the gel. The gel was then expanded in low salt buffer and finally embedded again in a non-expandable gel to stabilize it in the expanded state (Fig. 2A). The RNA was anchored to the gel at its poly(A) tail, which creates a single-pointed contact between individual RNA molecules and the gel and avoids stretching the mRNA during expansion process. This should facilitate separation of the nearby molecules. MERFISH encoding probes were stained before embedding samples in a gel since it was found that MERFISH probes did not penetrate reliably well after samples were embedded twice, at least in some embodiments. A low salt buffer (0.5x saline-sodium citrate (SSC)) was used instead of water for dialysis to preserve specific binding of probes during expansion, which resulted in a lower expansion factor. Measurements of the gel volume showed that dialysis in 0.5x saline-sodium citrate (SSC) produced a 12-fold expansion in volume. It was noted that as long as the expansion is sufficient to separate neighboring RNA molecules, a lower expansion factor had the advantage of allowing faster imaging, in certain applications.

Notably, in the expanded samples in these experiments, individual RNA molecules became well resolved (Fig. 2B, C) and were successfully decoded (Fig. 2D). The average copy number per cell detected for these RNAs by MERFISH correlated with the RNA abundance measured by RNA-seq with a Pearson correlation coefficient of 0.83 between the $\log_{10}$ values (Fig. 2E).

To quantify the improvement on decoding performance with expansion, the average MERFISH counts per RNA species per cell were compared between expanded and

unexpanded samples. It was found that the copy numbers per cell for the 129 RNA species detected in expanded samples correlated strongly with those in the unexpanded samples with a high Pearson correlation coefficient of 0.88 between the $\log_{10}$ values (Fig. 2F). However, the copy numbers per cell detected in expanded samples were 7.5 +/- 0.3 fold (average +/- SEM) or 6.6 fold (median) higher than those detected in unexpanded samples for these 129 RNA species, indicating that the detection efficiency for the expanded sample was substantially higher compared to the unexpanded sample for this high-abundance library. The copy number per cell results were highly reproducible between replicates of experiments (Fig.2G). Comparison with smFISH measurements showed that the copy numbers per cell determined by MERFISH was 105% +/- 9% (average +/- SEM) or 111% (median) of those determined by smFISH (Fig. 2H), indicating that the detection efficiency after expansion was close to 100%.

Fig. 2 shows MERFISH measurements of a high-abundance RNA library in expanded U-2 OS cells. Fig.2 A shows a schematic representation of the basic implementation of expansion MERFISH. Target RNAs were co-stained with acrydite-modified poly-dT anchoring probes and MERFISH encoding probes, and then anchored to an expandable polymer gel via the poly-dT probes. Afterwards, samples were incubated in digestion buffer with Proteinase K and SDS for removal of proteins and lipids respectively to homogenize mechanical properties of the gel. Finally, the gel was then expanded by dialysis in low salt buffer and the expanded gel was embedded polyacrylamide gel again to stabilize it in the expanded state. Fig. 2B shows an image of MERFISH measurement in a U-2 OS sample stained with encoding probes for 129 RNAs, embedded, cleared, expanded, re-embedded and visualized with a Cy5-labeled readout probe at one focal plane of the z-scan. The solid lines mark the edge of a cell and the dashed lines mark the DAPI-stained region of a nucleus. Fig. 2C shows a high-pass filtered fluorescence images of all 8 rounds of two-color smFISH imaging for the boxed sub-region in Fig. 2B. Shading represent the Cy5 channel (green) and Alexa 750 channel (red). Fig. 2D shows the localizations of all decoded RNAs in Fig. 2B colored according to their measured binary barcodes. The inset shows the localizations of all decoded RNAs from the region shown in Fig. 2C. Decoded RNAs across all optical sections are displayed. The solid lines mark the edge of a cell and the dashed lines mark the DAPI-stained region of a nucleus. Fig. 2E shows the average RNA copy numbers per cell for the 129 RNA species determined by MERFISH vs. the abundances as determined by RNA-seq. The Pearson correlation coefficient between the $\log_{10}$ values (ρ_{10}) is 0.83. Fig.2 F shows the average RNA copy numbers per cell for the 129 RNA species determined by expansion

MERFISH vs. those by MERFISH in unexpanded samples. A z-scan of 12 micrometers in depth was performed for expansion MERFISH measurements and a z-scan of 5 micrometers in depth was performed for unexpanded samples to ensure that at least 90% of the RNA molecules were included in the imaged depth. The Pearson correlation coefficient between the $\log_{10}$ values (ρ_{10}) is 0.88. The copy numbers per cell detected in expanded samples were 7.5 +/- 0.3 fold (average +/- SEM) or 6.6 fold (median) of those detected in unexpanded samples for these 129 RNA species. Fig. 2G shows the average RNA copy numbers per cell for the 129 RNA species detected in one expanded sample vs. a second one. The Pearson correlation coefficient between the log10 values (ρ_{10}) is 0.94. Fig. 2H shows the average RNA copy numbers per cell determined by expansion MERFISH vs. those by smFISH measurements of cleared and unexpanded samples for 12 of the 129 RNAs. smFISH measurements were performed for the unexpanded but clear sample as shown in Fig. 1. The Pearson correlation coefficient between the $\log_{10}$ values (ρ_{10}) is 0.96. The mean ratio between MERFISH and smFISH results was 105% +/- 9% (SEM, n = 12 RNA species) and the median ratio 111%.

EXAMPLE 4

Incorporating immunofluorescence imaging into MERFISH experiments of expanded samples. A cell is a highly complex system composed of different structures and compartments and immunofluorescence is a powerful technique to visualize specific subcellular structures and compartments. This example thus tested whether combination of immunofluorescence and MERFISH imaging was possible in the expanded samples. Expansion microscopy has been demonstrated in immunostained samples, using oligo-conjugated antibodies and complementary probes with a methacryloyl group to incorporate signals into the gel. These oligos then mark the positions of the protein target and can be detected by hybridization of fluorophore-labeled complementary oligos, which were incorporated into the MERFISH readout measurements. In the experimental procedure, immunostaining was performed of the protein targets with oligo-labeled (secondary) antibodies after hybridization of the cells with the MERFISH encoding probes and RNA anchor probes, and then the labeled sample was embedded in an expandable gel. Acrydite modification was added to the oligo on the antibodies so that it could be incorporated into the polymer gel during the embedding step. After digestion, gel expansion and second embedding in a non-expandable gel as described above, the MERFISH readout procedure was performed to first detect the readout sequences in the RNA encoding probes and then

with an additional round of FISH detection to read out the oligo sequences representing the protein target.

For this example, cadherin was immunostained in cultured U-2 OS cells, which were stained with the same high-abundance MERFISH library described above. Both specific staining of cadherin on the cell periphery (Fig. 3A) as well as clearly resolved smFISH spots in the same cells (Fig. 3B-D) were observed. The combination with immunofluorescence did not affect MERFISH imaging quality (Figs. 3E and 3F). The same high correlation of MERFISH results and the RNA-seq results with a Pearson correlation coefficient of 0.83 between the $\log_{10}$ values (Fig. 3E) was observed. The average copy numbers per RNA species per cell detected in immunostained samples correlated strongly with those in detected samples not subjected to immunostaining, with a Pearson correlation coefficient of 0.95 between the $\log_{10}$ values (Fig. 3F). On average, the ratio of copy numbers per RNA species per cell between immunostained and non-immunostained samples is 99% +/- 3% (SEM), indicating negligible impairment in performance of MERFISH when combined with immunofluorescence. The staining of cadherin in expanded and MERFISH labeled samples also looked similar to cadherin staining in control U-2 OS samples immuostained directly after fixation without MERFISH imaging of the RNA (Fig. 3G). To quantify whether antibody staining was affected by MERFISH RNA labeling, gel embedding, digestion, and expansion procedures, the cadherin intensity per unit length on the cell periphery (normalized to the actual length before expansion) was compared in expansion MERFISH samples versus control samples, with comparable results (Fig. 3H).

Fig. 3 shows incorporating immunofluorescence imaging into MERFISH measurements of expanded samples. Fig. 3A shows an image of an expanded U-2 OS sample stained with MERFISH encoding probes, cadherin primary antibodies, oligo-conjugated secondary antibodies and a Cy5-conjugated complementary probe. Fig. 3B shows smFISH image of MERFISH measurement for the boxed sub-region in Fig. 3A, visualized with a Cy5-labeled readout probe at one focal plane of the z-scan. Fig. 3C shows a high-pass filtered fluorescence images of all 8 rounds of two-color smFISH imaging for the boxed sub-region in Fig. 3B. Shading represent the Cy5 channel (green) and Alexa 750 channel (red). Fig. 3D shows the localizations of all decoded RNAs in Fig. 3B colored according to their measured binary barcodes. The inset shows the localizations of all decoded RNAs of the region shown in Fig. 3C. Decoded RNAs across all optical sections are displayed. Fig. 3E shows the average RNA copy numbers per cell for the 129 RNA species determined by MERFISH vs. the abundances as determined by RNA-seq. The Pearson correlation coefficient between the

$\log_{10}$ values (ρ_{10}) was 0.83. Fig. 3F shows the average RNA copy numbers per cell for the 129 RNA species determined by expansion MERFISH in immunostained samples vs. those in MERFISH-only samples. A z-scan of 12 micrometer in depth was performed for both. The Pearson correlation coefficient between the $\log_{10}$ values (ρ_{10}) was 0.95. The mean ratio between the RNA copy numbers determined in the two experiments was 99% +/- 3% (SEM, n = 129 RNA species) and the median ratio 94%. Fig. 3G shows an image of unexpanded cultured U-2 OS cells stained with cadherin primary antibodies, oligo-conjugated secondary antibodies and a Cy5-conjugated complementary probe right after fixation and permeabilization. Fig. 3H shows cadherin intensity per unit length (normalized to the actual length before expansion) in expanded MERFISH and control samples. Intensities were combined across stacks of z-scan after removal of background.

EXAMPLE 5

These examples have demonstrated an approach to combine MERFISH and expansion microscopy to measure high-abundance RNA libraries. It was shown that when the total RNA density is high, the overlap between signals from nearby RNA molecules reduced the MERFISH detection efficiency and that sample expansion could overcome this overlapping problem and substantially increase the detection efficiency, recovering the near 100% detection efficiency of MERFISH for these high-abundance RNA libraries.

Each mammalian cell has tens of thousands of different RNA species. MERFISH imaging can be used to determine thousands of RNA species in single cells. To increase the number of RNA species that can be simultaneously imaged in single cells while maintaining similar number of imaging rounds, the density of RNAs imaged per round may increase and may likely become an important limiting factor in the number of RNAs that can be imaged. Expansion MERFISH approach can facilitate a substantial increase in the number of RNA species that can be measured in single cells.

In these examples, the combination of MERFISH with immunofluorescence in these expanded samples was demonstrated, which can provide important information on the cellular context for transcriptome analysis. In conventional immunofluorescence with fluorophore-conjugated antibodies, the number of spectrally-distinct fluorophores limits the number of targets that can be studied simultaneously. The use of oligo-conjugated antibodies potentially allows visualizing many proteins in sequential rounds of hybridization and imaging using just one or a few spectrally distinct fluorophores. Thus, the combination of MERFISH and immunofluorescence imaging may allow a combined proteomic and transcriptomic measurements simultaneously in single cells.

EXAMPLE 6

This example describes various materials and methods used in some of the above examples.

Design of the encoding probes. The MERFISH-encoding probes were designed using 16-bit Hamming-weight-4 Hamming-distance-4 code with 140 possible barcodes. In this encoding scheme, all barcodes used were separated by a Hamming distance of at least 4, and hence at least four bits must be read incorrectly to change one valid barcode to another. Constant Hamming weight (i.e. the number of "1" bits in each barcodes) of 4 is used to avoid potential measurement bias due to differential rate of "1" to "0" and "0" to "1" errors. The encoding probe set contained 92 encoding probes per RNA. Each encoding probe was comprised of a 30-nt target region designed using a pipeline, flanked by two 20-nt readout sequences randomly selected out of the four ones assigned to each RNA, one 20-nt priming region at the 5' end and another 20-nt priming region using the reverse complement of the T7 promoter at the 3' end. Additional adenosine nucleotide spacers were added between readout sequences and target regions to prevent terminal guanine triplets in the readout sequences and to prevent target regions from combining with Gs from adjacent sequences to form G quadruplets. The priming region at the 5' end had a thymine at the end, which was put at the junction of the priming regions at the 5' end and the encoding region. This was designed to incorporate a uracil on the forward primer used in the reverse transcription step of probe construction, which can be cleaved by Uracil-Specific Excision Reagent (USER) Enzyme. The cleavable primer design together with using reverse complement of the T7 promoter as a second primer allowed construction of encoding probes with a length of 72 nt. The reduction in probe length may facilitate penetration of probes and reduce non-specific binding.

To include some abundant but short RNAs in our library, instead of designing probes targeting distinct regions of RNAs, the probes were allowed to share up to 20 nt with another probe. This modification helped increase the number of probes that could be designed for each gene by 2 fold. 92 target regions per RNA were selected randomly out of all potential target regions of an RNA. By allowing up to a 20-nt overlap between neighboring encoding probes, this design allowed RNAs as short as 1,200 nt to be targeted by 92 encoding probes with a 30-nt targeting sequence. Because a given cellular RNA is typically bound by less than one third of the 92 encoding probes, the encoding probes with overlapping targeting regions were expected to not substantially interfere with each other, but would partially compensate for reduced binding due to local inaccessible regions on the target RNA (e.g. secondary structure) or loss of probe during synthesis.

Construction of the encoding probes. The encoding probe set was constructed from complex oligonucleotide pools. Briefly, the oligopools (CustomArray) were amplified via limited-cycle PCR to make in vitro transcription templates, converted these templates into RNA via in vitro transcription (New England Biolabs), and converted the RNA back to DNA via reverse transcription (Maxima RT H, Thermo Fisher Scientific). The probes were then digested by USER Enzyme (New England Biolabs) at a dilution of 1:30 (vol/vol) incubated at 37 °C for 24 h to cleave off the priming region at the site of a uracil between a priming regions at the 5' end and the target region. After that, DNA was purified via alkaline hydrolysis to remove RNA and column purification (Zymo Research). The final probes were resuspended in RNAase-free water and stored at -20 °C.

Oligo conjugation to secondary antibodies. Oligonucleotides containing the desired readout probes were conjugated to secondary antibodies via a combination of NHS-ester and copper-free click chemistries. First, secondary antibodies were labeled with a copper-free click crosslinking agent using NHS-ester chemistry. Specifically, azide preservative was removed from the unconjugated Donkey Anti-Rabbit secondary antibodies (Thermoscientific) using a spin-column based dialysis membrane (Amicon, 100kDa molecular weight cut off) according to the manufacturer's instructions. The NHS-ester, PEG5, DBCO cross linker (Kerafast) was diluted to a concentration of 10 micromolar in anhydrous DMSO (Thermoscientific). 2 microliters of this cross linker was then combined with 100 microliters of 2 mg/mL of the antibody in 1x phosphate-buffered saline (PBS). This reaction was incubated at room temperature for 1 hour and then terminated via a second round of purification using the Amicon columns as described above. The average number of DBCO crosslinkers per antibody was determined via the relative absorption of the sample at 280 nm (antibody) and 309 nm (DBCO). On average the crosslinker amounts described above produced ~7 DBCO crosslinkers per antibody.

Oligonucleotide probes containing the desired sequence as well as a 5'-acrydite, to allow cross linking to the polymer gel, and a 3'-azide, to allow crosslinking to the DBCO-labeled antibodies, were ordered from IDT and suspended to 100 micromolar in 1x PBS. 20 microliters of the appropriate oligonucleotide was then added to 100 microliters of the DBCO-labeled antibodies at a final concentration of ~2 mg/mL. This reaction was incubated at 4 °C for at least 12 hours. Labeled antibodies were not further purified as residual oligonucleotides, not conjugated to antibodies, are expected to be readily washed away from samples.

Cell culture and fixation. U-2 OS cells (ATCC) were cultured with Eagle's Minimum Essential Medium (ATCC) containing 10% (vol/vol) FBS (Thermo Fisher Scientific). Cells were plated on 40-mm-diameter, no. 1.5 coverslips (Bioptechs) at 350,000 cells per coverslip and were incubated in Petri dishes at 37 °C with 5% CO_2 for 48 h. Cells were fixed, permeabilized, and stained with encoding probes. Briefly, cells were fixed for 15 min in 4% (vol/vol) paraformaldehyde (Electron Microscopy Sciences) in 1x PBS at room temperature, washed three times with 1x PBS, permeabilized for 10 min with 0.5% (vol/vol) Triton X-100 (Sigma) in 1x PBS at room temperature, and washed once with 1x PBS.

Encoding probe staining for MERFISH measurement. Permeabilized cells were incubated for 5 min in encoding wash buffer comprising 2x saline-sodium citrate (SSC) (Ambion) and 30% (vol/vol) formamide (Ambion). Then 30 microliters of ~300 micromolar encoding probes and 3.3 micromolar of anchor probe (a 20-nt sequence of alternating dT and thymidine-locked nucleic acid (dT+) with a 20-nt reverse complement of a readout sequence and a 5′-acrydite modification (Integrated DNA Technologies)) in encoding hybridization buffer was added to the surface of Parafilm (Bemis) and was covered with a cell-containing coverslip. Samples then were incubated in a humid chamber inside a hybridization oven at 37 °C for 40 h. Encoding hybridization buffer was composed of encoding wash buffer supplemented with 0.1% (wt/vol) yeast tRNA (Life Technologies), 1% (vol/vol) murine RNase inhibitor (New England Biolabs), and 10% (wt/vol) dextran sulfate (Sigma). Cells then were washed with encoding wash buffer and were incubated at 47 °C for 30 min; this washing step was repeated once.

Immunostaining. After being stained with encoding and anchor probes, samples were post-fixed with 4% (vol/vol) paraformaldehyde (Electron Microscopy Sciences) at room temperature for 10 min. Samples were then blocked at room temperature for 30 min in blocking buffer with 4% (wt/vol) UltraPure BSA (Thermo Fisher Scientific) in 2x SSC (Ambion) supplemented with 3% (vol/vol) RNasin Ribonuclease inhibitors (Promega), 6% (vol/vol) murine RNAase inhibitors (New England Biolabs) and 1 mg/ml yeast tRNA (Life Technologies). Samples were incubated with primary antibodies (anti-pan Cadherin, Abcam) in blocking buffer at a concentration of 2 micrograms/ml for 1 h at room temperature, and washed three times with 2x SSC for 10 min each. Samples were then incubated with oligo-labeled secondary antibodies in blocking buffer at a concentration of 3.75 micrograms/ml for 1 h at room temperature, then washed with 2x SSC three times for 10 min each. Samples were fixed again with 4% (vol/vol) paraformaldehyde (Electron Microscopy Sciences) for 10 min.

Embedding, digestion and clearing for unexpanded samples. Stained samples on coverslips were first incubated for 5 min with a de-gassed polyacrylamide solution with 4% (vol/vol) of 19:1 acrylamide/bis-acrylamide (BioRad), 60 mM Tris·HCl pH 8 (ThermoFisher), 0.3 M NaCl (ThermoFisher), 0.2% (vol/vol) Tetramethylethylenediamine (TEMED) and a 1:25,000 dilution of 0.1-micrometer-diameter carboxylate-modified orange fluorescent beads (Life Technologies). The beads served as fiducial markers for the alignment of images taken across multiple rounds of smFISH imaging.

To cast a thin polyacrylamide film, for each sample, 50 microliters of the final solution was added to the surface of a glass plate (TED Pella) that had been pretreated for 5 min with 1 mL GelSlick (Lonza) so as not to stick to the polyacrylamide gel. Samples on coverslips, treated as described above, were aspirated and gently inverted onto this 50-microliter droplet to form a thin layer of solution between the coverslip and the glass plate. The solution was then allowed to polymerize for 2 h at room temperature in a home-built chamber filled with nitrogen. The coverslip and the glass plate were then gently separated, and the PA film was washed once with a digestion buffer with 2% (wt/vol) SDS (Thermo Fisher Scientific), 0.5% (vol/vol) Triton X-100 in 2x SSC (Ambion). The digestion buffer used 2% (wt/vol) SDS to facilitate lipid removal. After the wash, the gel was covered with digestion buffer supplemented with 1% (vol/vol) Proteinase K (New England Biolabs). The sample was digested in this buffer for >12 h in a humidified, 37 ^{o}C incubator and then washed three times with 2x SSC. MERFISH measurements were either performed immediately or the sample was stored in 2x SSC supplemented with 0.1% (vol/vol) murine RNase inhibitor (New England Biolabs) at 4 ^{o}C for no longer than 48 h.

Silanization of coverslips for unexpanded samples. 40-mm-diameter No. 1.5 coverslips (Bioptechs) were washed for 30 min via immersion in a 1:1 mixture of 37% (vol/vol) HCl and methanol at room temperature. The coverslips were then rinsed three times in deionized water and once in 70% (vol/vol) ethanol. The coverslips were blown dry with nitrogen gas and then immersed in 0.1% (vol/vol) triethylamine (Millipore) and 0.2% (vol/vol) allyltrichlorosilane (Sigma) in chloroform for 30 min at room temperature. The coverslips were washed once each with chloroform and ethanol and then blown dry with nitrogen gas. Silanized coverslips were then stored at room temperature in a desiccated chamber overnight before use, to dehydrate the silane layer.

Embedding, digestion and clearing for expanded samples. The monomer solution containing 2 M NaCl (Thermo Fisher Scientific), 7.7% (wt/wt) sodium acrylate (Sigma), 4% (vol/vol) of 19:1 acrylamide/bis-acrylamide (BioRad) and 60 mM Tris HCl pH 8 (Thermo

Fisher Scientific) was prepared, frozen in aliquots at -20 °C, and thawed before use. The monomer solution was degassed and cooled to 4 °C before use. TEMED with a final concentration of 0.2% (vol/vol) and a 1:5,000 dilution of 0.1-micrometer-diameter carboxylate-modified orange fluorescent beads (Life Technologies) was added to the solution. The beads served as fiducial markers for the alignment of images taken across multiple rounds of smFISH imaging. Stained samples were incubated in the solution for 5 min at room temperature. The solution was then kept on ice and further supplemented with ammonium persulfate (Sigma) at a final concentration of 0.2% (wt/vol).

To cast a thin polymer film, for each sample, 50 microliters of the final solution was added to the surface of a glass plate (TED Pella) that had been pretreated for 5 min with 1 mL GelSlick (Lonza) so as not to stick to the gel. Samples on coverslips, treated as described above, were aspirated and gently inverted onto this 50-microliter droplet to form a thin layer of solution between the coverslip and the glass plate. The solution was then allowed to polymerize for 2 h at room temperature in a home-built chamber filled with nitrogen. The coverslip and the glass plate were then gently separated, and the gel was washed once with a digestion buffer with 2% (wt/vol) SDS (Thermo Fisher Scientific), 0.5% (vol/vol) Triton X-100 in 2x SSC (Ambion) as described above. After the wash, the thin film of the gel were carefully trimmed to desired sizes using a razor blade, and covered with digestion buffer supplemented with 1% (vol/vol) Proteinase K (New England Biolabs). The sample was digested in this buffer for >12 h in a humidified, 37 °C incubator and then washed with 2x SSC (Ambion) three times. The gel would expand ~1.5 fold during digestion.

Expansion and re-embedding. After digestion, samples were expanded in 0.5x SSC buffer supplemented with 0.2% (vol/vol) Proteinase K (New England Biolabs) at room temperature. Proteinase K was added to maintain samples in an RNAase free environment. The buffer was changed every 30 min until samples no longer expanded (typically about 2 h). Expanded gels were re-embeded in polyacrylamide gel to stabilize the gel for sequential rounds of readout probe hybridization and imaging. Briefly, samples were incubated in re-embedding solution composed of 4% (vol/vol) of 19:1 acrylamide/bis-acrylamide (BioRad)) with 30 mM NaCl (Thermo Fisher Scientific), 6 mM Tris HCl pH 8 (Thermo Fisher Scientific) and 0.2% (vol/vol) of TEMED (Sigma) for 20 min at room temperature on a rocker. The re-embedding solution was then kept on ice and further supplemented with ammonium persulfate (Sigma) at a final concentration of 0.2% (wt/vol). Gels were placed on a bind-silane-treated coverslip (see below), rinsed with the solution and dried quickly with

KimWipes (Kimtech). Coverslips with gels were put in a home-built nitrogen chamber, covered a glass plate (TED Pella) and allowed to polymerize at room temperature for 1 h.

Bind-silane treatment of coverslips. 40-mm-diameter, no. 1.5 coverslips (Bioptechs) were sonicated in 1 M potassium hydroxide for 30 min, wash three times with deionized water and sonicated again in 70% (vol/vol) ethanol for 30 min. The coverslips were then immersed in 5% (vol/vol) glacial acetic acid and 0.38% (vol/vol) bind-silane (GE Healthcare) in 99% (vol/vol) ethanol for 1 h at room temperature. After being quickly washed with 70% (vol/vol) ethanol three times, the coverslips were put into a 60 ^{o}C oven until dried completely. Coverslips can be stored in seal containers with desiccants for up to a month.

Sequential rounds of readout probe staining and imaging. To aid choosing the right focal plane for imaging, embedded samples were hybridized in dish with the first pair of two-color readout probes in hybridization buffer composed of 2x SSC (Ambion), 5% (vol/vol) ethylene carbonate (Sigma), 0.1% (vol/vol) murine RNase inhibitor in nuclease-free water, and 3 nM of the appropriate readout probes, for 30 min (expanded samples) or 10 min (unexpanded samples) at room temperature and washed for 20 min (expanded samples) or 7 min (unexpanded samples) in wash buffer composed of 2x SSC (Ambion) and 10% (vol/vol) ethylene carbonate (Sigma) in nuclease-free water. Samples were then washed with 2x SSC once, stained with DAPI at 10 micrograms/microliter in 2x SSC (Ambion) for 10 min, and washed 3 times in 2x SSC (Ambion) for 5 min each.

Samples were mounted into a flow chamber and the buffer exchange through this chamber for following imaging and hybridization was controlled via a home-built fluidics system composed of three computer-controlled eight-way valves (Hamilton) and a computer-controlled peristaltic pump (Gilson), with elongated flow and incubation time for expanded samples (besides the hybridization and wash time as described in the previous section. Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) incubation time was increased to 30 min and buffer exchange time to 7 min) to allow diffusion to reach equilibrium inside of the gel.

In order to accurately compare the counts per cell between different samples, more than 90% of target RNAs in a cell should be caught. To quantify RNA distribution across different z planes, optical sectioning was performed at discrete imaging planes across the cell at a step size of 0.5 micrometers and the distribution of smFISH signals was quantified as a function of z position. It was found that >90% of RNAs located in the first 5 micrometer-depth volume in unexpanded samples and in the first 12 micrometer-depth volume in expanded samples from the surface of coverslips. Thus, a 5 micrometer-depth volume was

scanned for unexpanded samples and a 12 micrometer-depth volume was scanned for expanded samples with a step size of 1 micrometer. The scanning in z direction was controlled by a Nano-F200 nanopositioner (Mad City Labs).

Sequential MERFISH imaging and signal removal was carried out on a high-throughput imaging platform. Briefly, after hybridization, samples were imaged with a FOV area of 223 x 223 micrometers utilizing a 2,048 x 2,048 pixel, scientific complementary metal-oxide semiconductor (sCMOS) camera in combination with a high numerical aperture (NA = 1.3) and a high-magnification (60x) silicone oil objective. After imaging ~100-400 FOVs, the fluorescence of the readout probes was extinguished by incubating the sample in a reductive cleavage buffer composed of 2x SSC and 50 mM TCEP (Sigma). The hybridization, imaging and chemical cleavage process was repeated eight times with 405-nm DAPI channel imaged in conjunction with the first round of readout imaging.

Image Registration and Decoding. Registration of images of the same FOV across imaging rounds as well as decoding of the RNA barcodes was conducted as follows. Briefly, z-stacks at each location from different imaging rounds were corrected for lateral offsets based on the location of fiducial beads. Corrected z-stacks were high-pass filtered to remove background, deconvolved to tighten RNA spots, and then low-pass filtered to connect RNA centroids that differed slightly in location between images. To correct for differences in the brightness between color channels, images were first normalized by equalizing their intensity histograms and refined further via an iterative process to remove substantial variation in the fluorescence intensity between different bits. The set of 16 normalized intensity values (corresponding to 16 readout probe imaging) observed for each pixel in each FOV at each focal plane represented a vector in a 16-dimensional space. The pixel vector was normalized and compared to each of the 140 barcodes in the 16-bit MHD4 code. A pixel was assigned to a barcode if the Euclidean distance between the vector and a barcode was smaller than a given threshold defined by the distance of a single-bit error. Adjacent pixels were combined into a single putative RNA using a 3-D connectivity array with maximal neighborhood connectivity. Cell boundaries were calculated using the watershed algorithm based on the inverted barcode density with DAPI- stained regions as initial seeds.

Computations were split between the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University and a desktop server that contained two 10-core Intel Xeon E5-2680 2.8GHz CPUs and 256 GB of RAM.

Single-molecule FISH. Each smFISH probe contains a 30-nt target region and a 20-nt custom-designed readout sequence. 48 probes were designed for each gene. After

permeabilization, cells were incubated for 5 min in encoding wash buffer comprising 2x saline-sodium citrate (SSC) (Ambion) and 30% (vol/vol) formamide (Ambion). Then 30 microliters of 2 micromolar smFISH probes and 3.3 micromolar of anchor probe in encoding hybridization buffer was added to the surface of Parafilm (Bemis) and was covered with a cell-containing coverslip. Samples then were incubated in a humid chamber inside a hybridization oven at 37 °C for 24 h. Encoding hybridization buffer was composed of encoding wash buffer supplemented with 0.1% (wt/vol) yeast tRNA (Life Technologies), 1% (vol/vol) murine RNase inhibitor (New England Biolabs), and 10% (wt/vol) dextran sulfate (Sigma). Cells then were washed with encoding wash buffer and were incubated at 47 °C for 30 min; this washing step was repeated once. Cells were then embedded and cleared using the matrix-imprinting-based clearing method described as described above for embedding, digestion and clearing for unexpanded samples. Embedded samples were hybridized in dish with a Cy5 20-nt readout probes in hybridization buffer, washed and stained with DAPI as described in Sequential rounds of readout probe staining and imaging Section. Images were collected scanning of a 5 micrometer-depth volume at a step size of 1 micrometer and 100 FOVs were collected for each smFISH probe set.

While several embodiments of the present invention have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the functions and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the present invention. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the teachings of the present invention is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, the invention may be practiced otherwise than as specifically described and claimed. The present invention is directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles,

materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the scope of the present invention.

In cases where the present specification and a document incorporated by reference include conflicting and/or inconsistent disclosure, the present specification shall control. If two or more documents incorporated by reference include conflicting and/or inconsistent disclosure with respect to each other, then the document having the later effective date shall control.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

The indefinite articles "a" and "an," as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean "at least one."

The phrase "and/or," as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with "and/or" should be construed in the same fashion, i.e., "one or more" of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the "and/or" clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to "A and/or B", when used in conjunction with open-ended language such as "comprising" can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, "or" should be understood to have the same meaning as "and/or" as defined above. For example, when separating items in a list, "or" or "and/or" shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as "only one of" or "exactly one of," or, when used in the claims, "consisting of," will refer to the inclusion of exactly one element of a number or list of elements. In general, the term "or" as used herein shall only be interpreted as indicating exclusive alternatives (i.e. "one or the other but not both") when preceded by terms of exclusivity, such as "either," "one of," "only one of," or "exactly one of."

As used herein in the specification and in the claims, the phrase "at least one," in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase "at least one" refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, "at least one of A and B" (or, equivalently, "at least one of A or B," or, equivalently "at least one of A and/or B") can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

When the word "about" is used herein in reference to a number, it should be understood that still another embodiment of the invention includes that number not modified by the presence of the word "about."

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

In the claims, as well as in the specification above, all transitional phrases such as "comprising," "including," "carrying," "having," "containing," "involving," "holding," "composed of," and the like are to be understood to be open-ended, i.e., to mean including but not limited to. Only the transitional phrases "consisting of" and "consisting essentially of" shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

CLAIMS

What is claimed is:

1. A method, comprising:

embedding cells within an expandable material;

immobilizing nucleic acids from the cells to the expandable material;

expanding the expandable material;

exposing the expandable material to a plurality of nucleic acid probes; and

determining binding of the nucleic acid probes to the immobilized nucleic acids.

2. The method of claim 1, wherein immobilizing the nucleic acids from the cells to the expandable material comprises immobilizing the nucleic acids at a single point to the expandable material.

3. The method of any one of claims 1 or 2, comprising immobilizing the nucleic acids to the expandable material.

4. The method of claim 3, comprising immobilizing the nucleic acids at a 5' end of the nucleic acids to the expandable material.

5. The method of any one of claims 3 or 4, comprising immobilizing the nucleic acids at a 3' end of the nucleic acids to the expandable material.

6. The method of any one of claims 3-5, comprising immobilizing the nucleic acids at a central portion of the nucleic acids to the expandable material.

7. The method of any one of claims 2-6, comprising immobilizing the nucleic acids at more than one location.

8. The method of any one of claims 1-7, wherein immobilizing the nucleic acids from the cells to the expandable material comprises exposing the nucleic acids to an anchor probe that immobilizes the nucleic acid relative to the expandable material.

9. The method of claim 8, wherein the anchor probe comprises a first portion that reacts with the nucleic acids from the cells and a second portion that reacts with the expandable material.

10. The method of claim 9, wherein a unique anchor probe is used for each RNA.

11. The method of claim 9, wherein one or more unique anchor probes are used for each RNA.

12. The method of any one of claims 9-11, wherein the first portion comprises a nucleic acid sequence substantially complementary with the nucleic acids from the cells.

13. The method of any one of claims 9-12, wherein the anchor probe comprises an acrydite-modified oligonucleotide.

14. The method of any one of claims 9-13, wherein the anchor probes comprises a poly-dT sequence.

15. The method of any one of claims 1-14, wherein embedding cells within an expandable material comprises surrounding the cells with a precursor of the expandable material, and causing the precursor of the expandable material to form the expandable material surrounding the cells.

16. The method of claim 15, wherein the precursor of the expandable material comprises a monomer.

17. The method of any one of claims 15 or 16, wherein causing the precursor to form the expandable material comprises polymerizing the precursor.

18. The method of any one of claims 15-17, wherein causing the precursor to form the expandable material comprises exposing the precursor to a cross-linking agent.

19. The method of any one of claims 1-18, wherein the precursor comprises acrylamide.

20. The method of any one of claims 1-19, wherein the precursor comprises an acrylate.

21. The method of any one of claims 1-20, wherein the expandable material comprises a gel.

22. The method of any one of claims 1-21, wherein the expandable material comprises polyacrylamide.

23. The method of any one of claims 1-22, wherein the expandable material comprises a methacrylate.

24. The method of any one of claims 1-23, comprising expanding the expandable material isotopically.

25. The method of any one of claims 1-24, further comprising fixing the cells.

26. The method of claim 25, wherein fixing the cells comprises exposing the cells to paraformaldehyde.

27. The method of any one of claims 1-26, wherein the nucleic acids from the cells comprises RNA.

28. The method of any one of claims 1-27, wherein the nucleic acids from the cells comprises mRNA.

29. The method of any one of claims 1-28, wherein the nucleic acids from the cells comprises DNA.

30. The method of any one of claims 1-29, wherein expanding the expandable material comprises exposing the expandable material to a solution comprising water.

31. The method of any one of claims 1-30, wherein expanding the expandable material comprises exposing the expandable material to a solution hypotonic to the expandable material.

32. The method of any one of claims 1-31, further comprising, for each at least some of the plurality of nucleic acid probes, determining binding of the nucleic acid probes within the sample.

33. The method of any one of claims 1-32, further comprising creating codewords based on the binding of the plurality of nucleic acid probes; and for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

34. The method of any one of claims 1-33, comprising exposing the cells to at least 5 different nucleic acid probes.

35. The method of any one of claims 1-34, comprising exposing the cells to the plurality of nucleic acid probes simultaneously.

36. The method of any one of claims 1-34, comprising exposing the cells to the plurality of nucleic acid probes sequentially.

37. The method of any one of claims 1-36, wherein the plurality of nucleic acid probes comprises a combinatorial combination of nucleic acid probes with different sequences.

38. The method of any one of claims 1-37, wherein at least some of the plurality of nucleic acid probes comprises a first portion comprising a target sequence and a second portion comprising one or more read sequences.

39. The method of claim 38, wherein the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from the one or more read sequences.

40. The method of any one of claims 38 or 39, comprising exposing the cells to a first secondary probe comprising a first signaling entity, the first secondary probe able to

bind to some of the read sequences of the nucleic acid probes, and determining binding of the nucleic acid probes by determining the first signaling entity.

41. The method of claim 40, further comprising exposing the cells to a second secondary probe comprising a second signaling entity, the second secondary probe able to bind to some of read sequences of the nucleic acid probes, and determining binding of the nucleic acid probes by determining the second signaling entity.

42. The method of any one of claims 1-41, wherein at least some of the plurality of nucleic acid probes comprise DNA.

43. The method of any one of claims 1-42, wherein at least some of the plurality of nucleic acid probes comprise RNA.

44. The method of any one of claims 1-43, wherein the plurality of nucleic acid probes have an average length of between 10 and 300 nucleotides.

45. The method of any one of claims 1-44, wherein at least some of the plurality of nucleic acids probes are configured to bind to the nucleic acids from the cells.

46. The method of any one of claims 1-45, wherein at least some of the binding of the nucleic acid probes to a target within the cells is specific binding.

47. The method of any one of claims 1-46, wherein at least some of the plurality of nucleic acids probes are configured to bind to RNA.

48. The method of any one of claims 1-47, wherein at least some of the plurality of nucleic acids probes are configured to bind to DNA.

49. The method of any one of claims 1-48, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 300 nm after expansion of the expandable material.

50. The method of any one of claims 1-49, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 100 nm after expansion of the expandable material.

51. The method of any one of claims 1-50, comprising determining binding of the nucleic acid probes within the sample at an effective resolution better than 30 nm prior to expansion of the expandable material.

52. The method of any one of claims 1-51, comprising determining binding of the nucleic acid probes within the sample at an effective resolution better than 10 nm prior to expansion of the expandable material.

53. The method of any one of claims 1-52, comprising determining binding of the nucleic acid probes by imaging at least a portion of the expandable material.

54. The method of any one of claims 1-53, comprising determining binding of the nucleic acid probes using an optical imaging technique.

55. The method of any one of claims 1-54, comprising determining binding of the nucleic acid probes using a fluorescence imaging technique.

56. The method of any one of claims 1-55, comprising determining binding of the nucleic acid probes using a super-resolution fluorescence imaging technique.

57. The method of any one of claims 1-56, further comprising determining genotypes of the cells by determining the binding of the nucleic acid probes.

58. A method, comprising:

immobilizing a plurality of targets to an expandable material, wherein at least 50% of the plurality of targets immobilized to the expandable material are immobilized at single points; and

expanding the expandable material.

59. A method, comprising:

immobilizing a plurality of nucleic acids to an expandable material, wherein at least 50% of the plurality of nucleic acids immobilized to the expandable material are immobilized at single points;

exposing the expandable material to a plurality of nucleic acid probes;

expanding the expandable material; and

determining binding of the nucleic acid probes to the immobilized nucleic acids.

60. An article, comprising:

an expandable material, comprising an embedded cell and a plurality of nucleic acids immobilized to the expandable material, wherein at least 50% of the plurality of nucleic acids immobilized to the expandable material are immobilized at single points.

61. An article, comprising:

a polymer comprising an expanded cell and a plurality of nucleic acids immobilized to the polymer, wherein the cell is expanded to at least 5 times its normal size within the polymer, and wherein at least 50% of the plurality of nucleic acids immobilized to the expandable material are immobilized at single points.

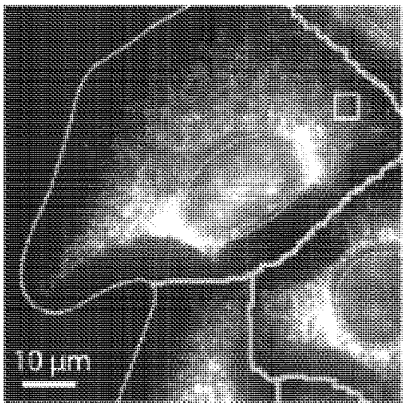


FIG. 1A

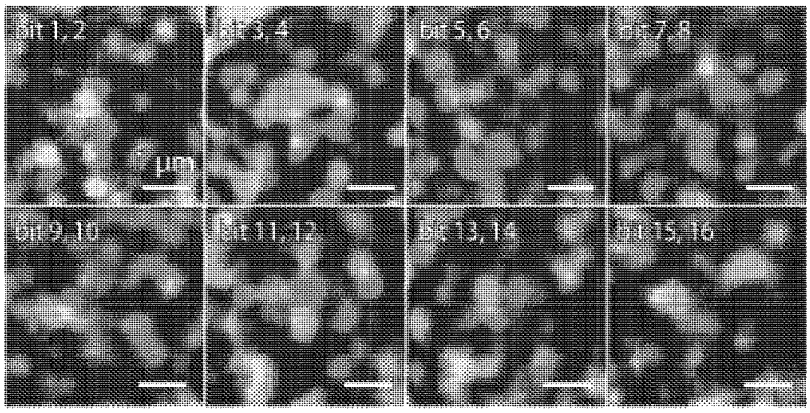


FIG. 1B

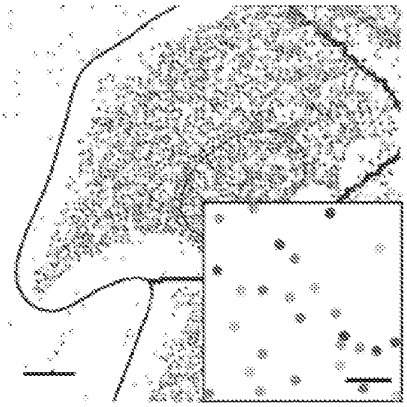


FIG. 1C

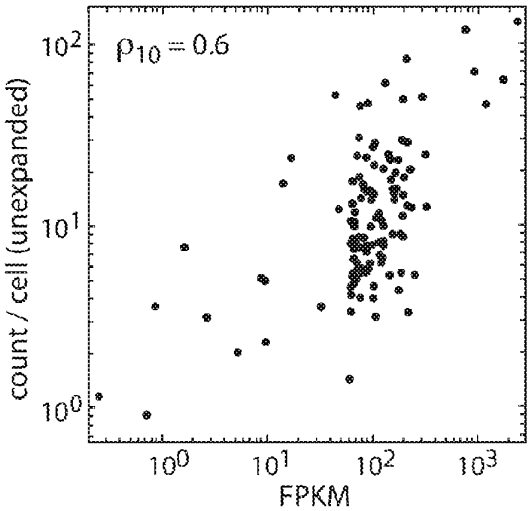

$\rho_{10} = 0.6$
count / cell (unexpanded)
FPKM
10^0
10^1
10^2
10^3

FIG. 1D

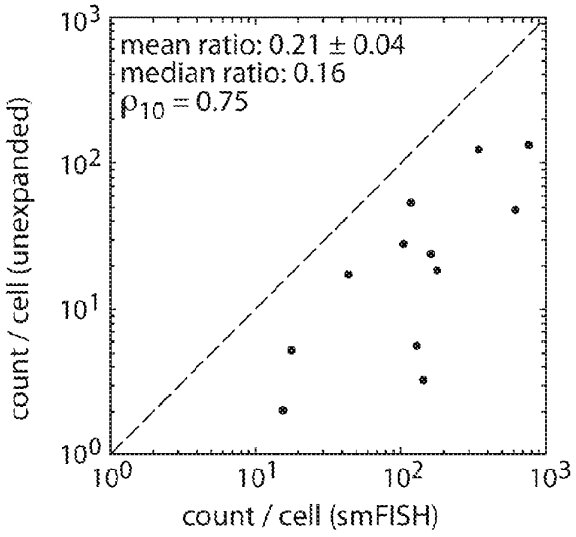

mean ratio: 0.21 ± 0.04
median ratio: 0.16
$\rho_{10} = 0.75$
count / cell (unexpanded)
count / cell (smFISH)
10^0
10^1
10^2
10^3

FIG. 1E

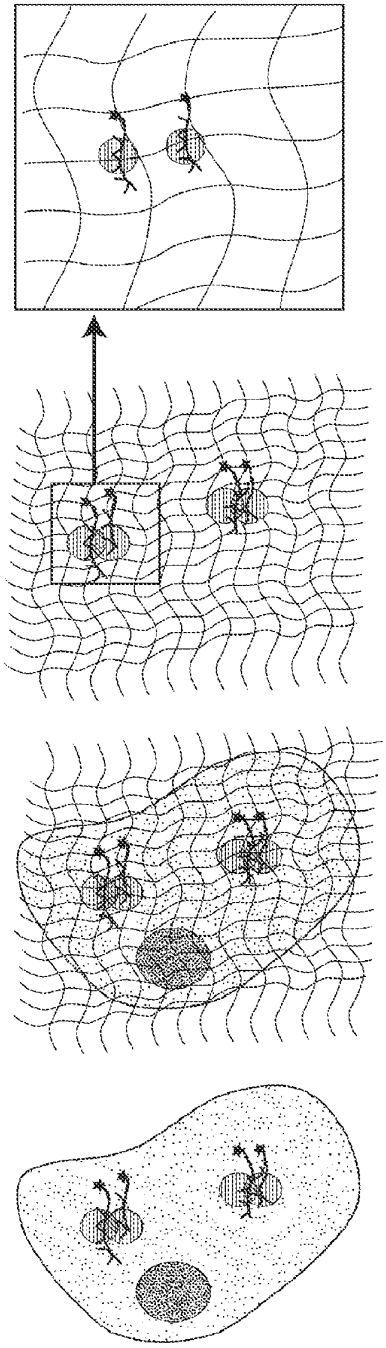


FIG. 2A

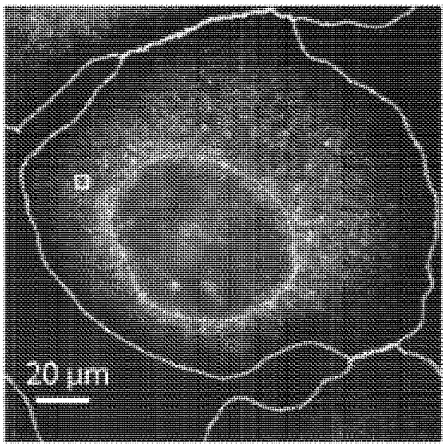

20 µm

FIG. 2B

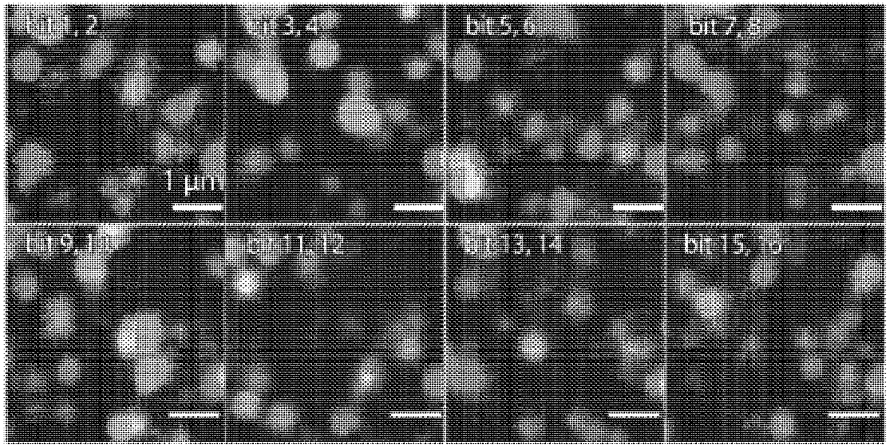

bit 2
bit 3, 4
bit 5, 6
bit 7, 8
1 µm
bit 9, 10
bit 11, 12
bit 13, 14
bit 15, 16

FIG. 2C

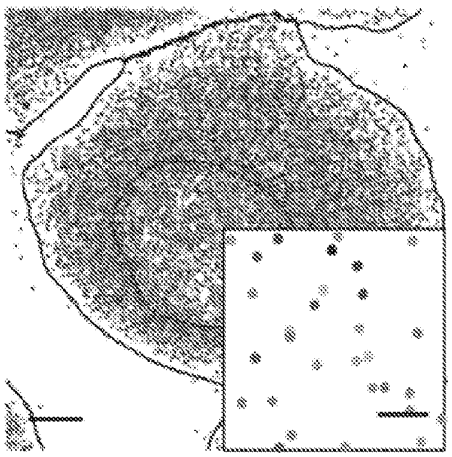


FIG. 2D

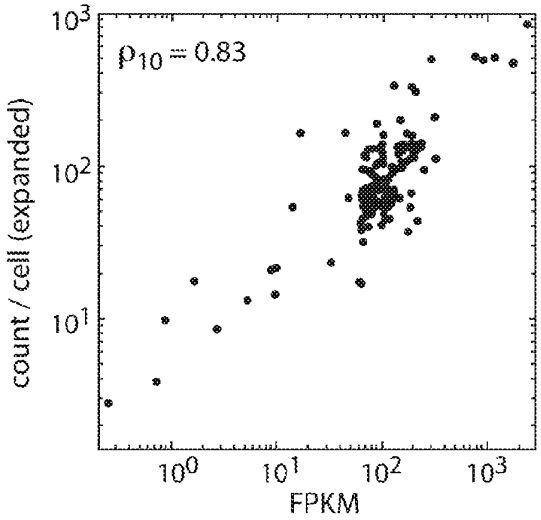

10^3
$\rho_{10} = 0.83$
10^2
10^1
count / cell (expanded)
10^0
10^1
10^2
10^3
FPKM

FIG. 2E

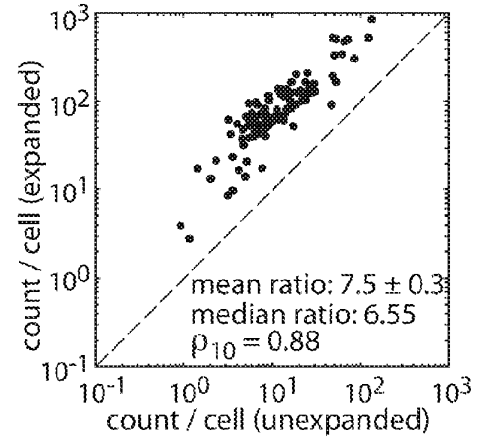


FIG. 2F

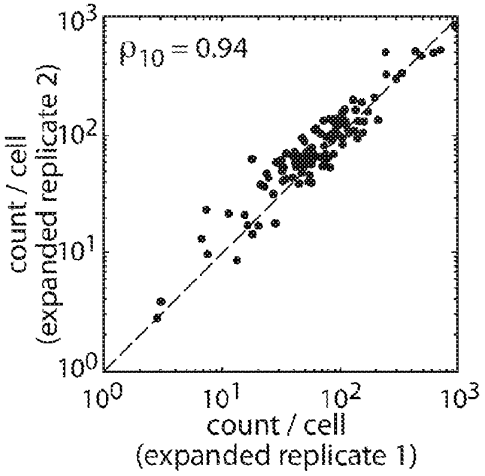


FIG. 2G

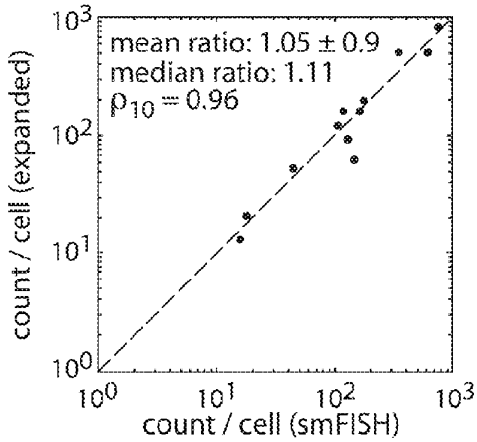


FIG. 2H

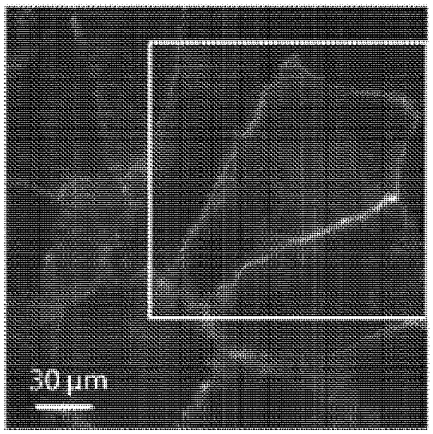


FIG. 3A

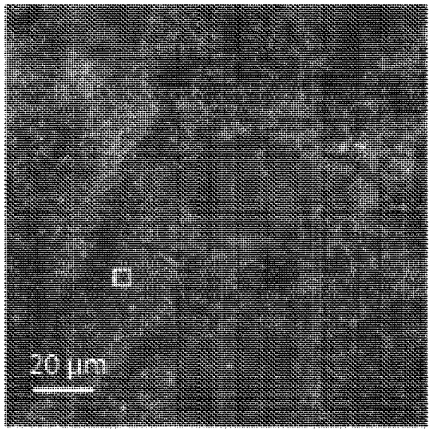


FIG. 3B

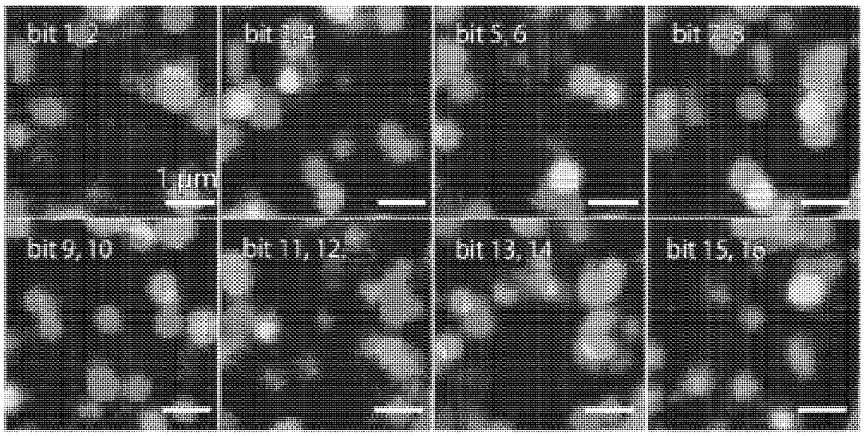


FIG. 3C

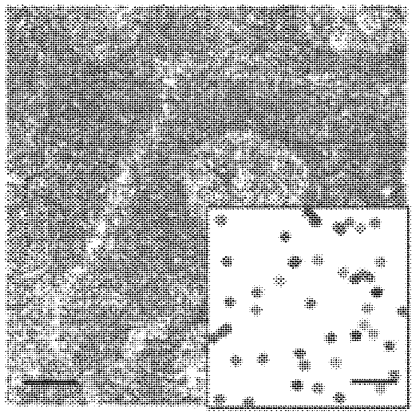


FIG. 3D

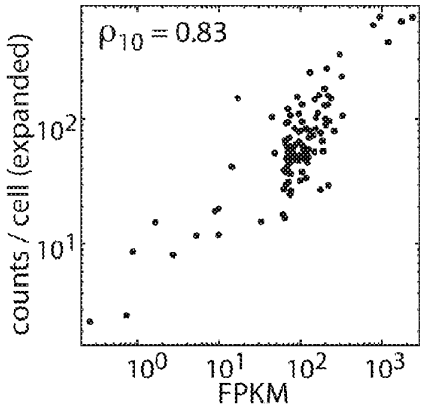


FIG. 3E

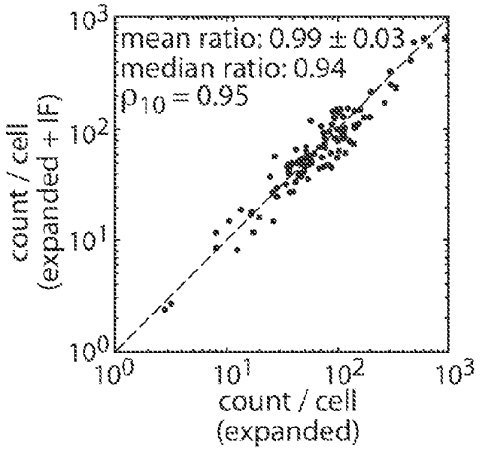


FIG. 3F

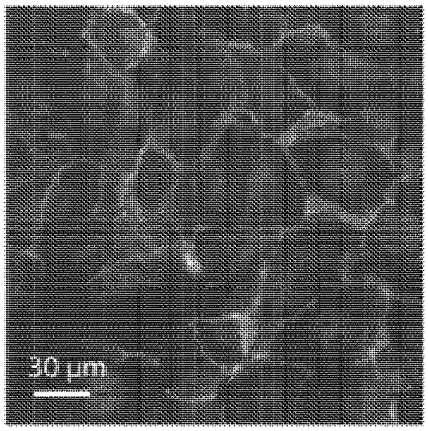


FIG. 3G

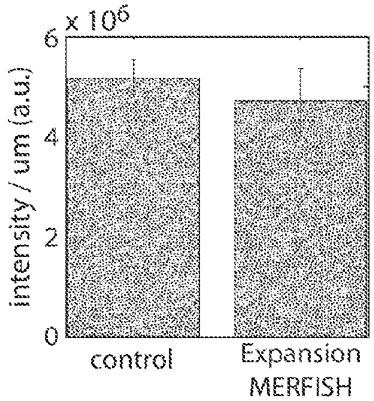


FIG. 3H

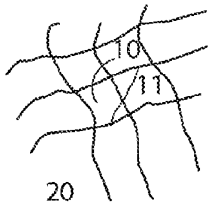


FIG. 4A

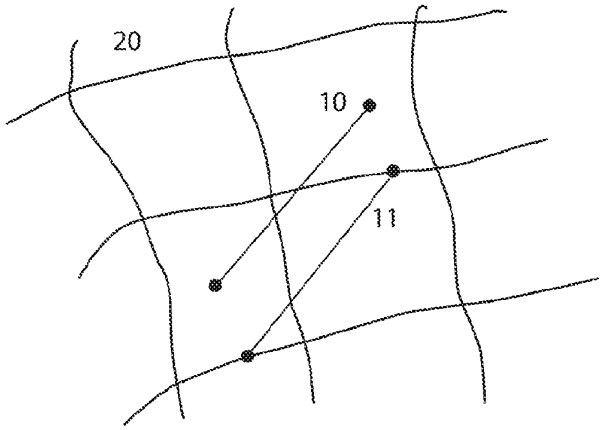


FIG. 4B

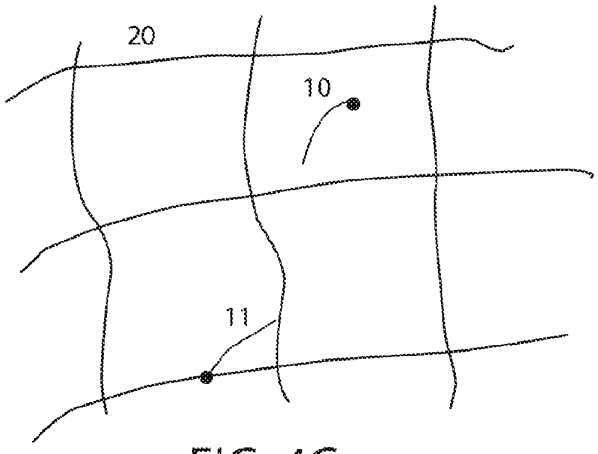


FIG. 4C

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 17/60558

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G01N 1/36, C12Q 1/68 (2017.01)
CPC - C12Q 1/6841, C12Q 1/6806, G01N 2001/364, G01N 1/36

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History Document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History Document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History Document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2016/0304952 A1 (Massachusetts Institute Of Technology et al.) 20 October 2016 (20.10.2016) para [0015]-[0022]; [0027]; [0029]; [0037]; [0049]	1-4, 58-61
X	Chen et al. "Nanoscale imaging of RNA with expansion microscopy" Nature Methods, 2016; 13:679-684. Published online 04 July 2016 (04.07.2016) (doi: 10.1038/nmeth.3899) whole document	1-4, 58-61
A	US 2016/0116384 A1 (Massachusetts Institute Of Technology) 28 April 2016 (28.04.2016) whole document	1-4, 58-61
A	Tillberg et al. "Protein-retention expansion microscopy of cells and tissues labeled using standard fluorescent proteins and antibodies" Nat Biotechnol. 2016; 34(9):987-992. Published online 04 July 2016 (04.07.2016) (doi: 10.1038/nbt.3625.) whole document	1-4, 58-61
A	Chen et al. "Expansion microscopy" Science, 30 Jan 2015 (30.01.2015): 347(6221):543-548 (doi: 10.1126/science.1260088) whole document	1-4, 58-61

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
04 January 2018	05 FEB 2018
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 17/60558

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 5-57
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (2)) (January 2015)

RECEIVED
Wolf, Greenfield & Sacks, P.C.

PATENT COOPERATION TREATY

IFD JUL 20 2020

From the INTERNATIONAL SEARCHING AUTHORITY

Docketed ✓ Already Docketed
Not Required
Initials 1st eg 2nd

To: Tani Chen
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT AND THE WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY, OR THE DECLARATION

(PCT Rule 44.1)

	Date of mailing *(day/month/year)* 16 JUL 2020
Applicant's or agent's file reference H0498.70693WO00	**FOR FURTHER ACTION** See paragraphs 1 and 4 below
International application No. PCT/US20/28632	International filing date *(day/month/year)* 17 April 2020 (17.04.2020)
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE	

1. ☒ The applicant is hereby notified that the international search report and the written opinion of the International Searching Authority have been established and are transmitted herewith.

   Filing of amendments and statement under Article 19:
   The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46):

   When? The time limit for filing such amendments is normally two months from the date of transmittal of the international search report.

   How? Directly to the International Bureau preferably through ePCT, or on paper to:
   The International Bureau of WIPO, 34, chemin des Colombettes, 1211 Geneva 20, Switzerland

   For more detailed instructions, see the *PCT Applicant's Guide*, International Phase, paragraphs 9.004 – 9.011.

2. ☐ The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect and the written opinion of the International Searching Authority are transmitted herewith.

3. ☐ **With regard to any protest** against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:
   - ☐ the protest together with the decision thereon has been transmitted to the International Bureau together with any request to forward the texts of both the protest and the decision thereon to the designated Offices.
   - ☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Reminders**

   The applicant may **submit comments on an informal basis on the written opinion of the International Searching Authority** to the International Bureau. These comments will be made available to the public after international publication. The International Bureau will send a copy of such comments to all designated Offices unless an international preliminary examination report has been or is to be established.

   Shortly after the expiration of **18 months from the priority date, the international application will be published** by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau before the completion of the technical preparations for international publication (Rules 90*bis*.1 and 90*bis*.3).

   Within **19 months** from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase **until 30 months** from the priority date (in some Offices even later); otherwise, the applicant must, **within 20 months** from the priority date, perform the prescribed acts for **entry into the national phase** before those designated Offices. In respect of other designated Offices, the time limit of **30 months** (or later) will apply even if no demand is filed within 19 months. For details about the applicable time limits, Office by Office, see www.wipo.int/pct/en/texts/time_limits.html and the *PCT Applicant's Guide*, National Chapters.

   Within **22 months from the priority date, the applicant may request that a supplementary international search be carried out** by a different International Searching Authority that offers this service (Rule 45*bis*.1). The procedure for requesting supplementary international search is described in the *PCT Applicant's Guide*, International Phase, paragraphs 8.006-8.032.

Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Shane Thomas Telephone No. PCT Helpdesk: 571-272-4300

Form PCT/ISA/220 (revised January 2020)

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

To: Tani Chen
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT AND THE WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY, OR THE DECLARATION

(PCT Rule 44.1)

	Date of mailing *(day/month/year)*
Applicant's or agent's file reference H0498.70693WO00	**FOR FURTHER ACTION** See paragraphs 1 and 4 below
International application No. PCT/US20/28632	International filing date *(day/month/year)* 17 April 2020 (17.04.2020)
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE	

1. ☒ The applicant is hereby notified that the international search report and the written opinion of the International Searching Authority have been established and are transmitted herewith.

   Filing of amendments and statement under Article 19:
   The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46):

   When? The time limit for filing such amendments is normally two months from the date of transmittal of the international search report.

   How? Directly to the International Bureau preferably through ePCT, or on paper to:
   The International Bureau of WIPO, 34, chemin des Colombettes, 1211 Geneva 20, Switzerland

   For more detailed instructions, see the *PCT Applicant's Guide*, International Phase, paragraphs 9.004 – 9.011.

2. ☐ The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect and the written opinion of the International Searching Authority are transmitted herewith.

3. ☐ **With regard to any protest** against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:
   ☐ the protest together with the decision thereon has been transmitted to the International Bureau together with any request to forward the texts of both the protest and the decision thereon to the designated Offices.
   ☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Reminders**

   The applicant may **submit comments on an informal basis on the written opinion of the International Searching Authority** to the International Bureau. These comments will be made available to the public after international publication. The International Bureau will send a copy of such comments to all designated Offices unless an international preliminary examination report has been or is to be established.

   Shortly after the expiration of **18 months from the priority date, the international application will be published** by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau before the completion of the technical preparations for international publication (Rules 90*bis*.1 and 90*bis*.3).

   Within **19 months** from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase **until 30 months** from the priority date (in some Offices even later); otherwise, the applicant must, **within 20 months** from the priority date, perform the prescribed acts for **entry into the national phase** before those designated Offices. In respect of other designated Offices, the time limit of **30 months** (or later) will apply even if no demand is filed within 19 months. For details about the applicable time limits, Office by Office, see www.wipo.int/pct/en/texts/time_limits.html and the *PCT Applicant's Guide*, National Chapters.

   Within **22 months from the priority date, the applicant may request that a supplementary international search be carried out** by a different International Searching Authority that offers this service (Rule 45*bis*.1). The procedure for requesting supplementary international search is described in the *PCT Applicant's Guide*, International Phase, paragraphs 8.006-8.032.

Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Shane Thomas Telephone No. PCT Helpdesk: 571-272-4300

Form PCT/ISA/220 (revised January 2020)

RECEIVED
Wolf, Greenfield & Sacks, P.C.
JUL 20 2020
Docketed Already Docketed
Not Required
Initials 1st 2nd

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference		FOR FURTHER ACTION
H0498.70693WO00		see Form PCT/ISA/220 as well as, where applicable, item 5 below.
International application No.	International filing date *(day/month/year)*	(Earliest) Priority Date *(day/month/year)*
PCT/US20/28632	17 April 2020 (17.04.2020)	19 April 2019 (19.04.2019)
Applicant		
PRESIDENT AND FELLOWS OF HARVARD COLLEGE		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 4 sheets.

☐ It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

a. With regard to the **language**, the international search was carried out on the basis of:

☒ the international application in the language in which it was filed.

☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

b. ☐ This international search report has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43.6*bis*(a)).

c. ☒ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, see Box No. I.

2. ☒ **Certain claims were found unsearchable** (see Box No. II).

3. ☐ **Unity of invention is lacking** (see Box No. III).

4. With regard to the **title**,

☒ the text is approved as submitted by the applicant.

☐ the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

☒ the text is approved as submitted by the applicant.

☐ the text has been established, according to Rule 38.2, by this Authority as it appears in Box No. IV. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. With regard to the **drawings**,

a. the figure of the **drawings** to be published with the abstract is Figure No. 1A

☒ as suggested by the applicant.

☐ as selected by this Authority, because the applicant failed to suggest a figure.

☐ as selected by this Authority, because this figure better characterizes the invention.

b. ☐ none of the figures is to be published with the abstract.

Form PCT/ISA/210 (first sheet) (July 2019)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US20/28632

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. ☒ forming part of the international application as filed:

      ☒ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US20/28632

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 4-75, 81-120, 124-134, 139-179
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US20/28632

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C12Q 1/6832, 1/68; C12N 9/22, 15/11 (2020.01)

CPC - C12N 9/22, 15/11; C12Q 1/6832, 1/6874

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Chen et al. 'Spatially Resolved, Highly Multiplexed RNA Profiling in Single Cells' Science, 24 April 2015; Vol 348, Issue 6233; pp 412, aaa6090-1-14; abstract; page 412, column 2, second paragraph; column 3, second paragraph; figure 1; figure 1 legend http://dx.doi.org/10.1126/science.aaa6090	1-2, 3/1-2, 76
Y	US 2018/0320226 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 08 November 2018; paragraphs [0004], [0009], [0015], [0027], [0072], [0167], [0169], [0190], [0191], [0193], [0234]	1-2, 3/1-2, 76-78, 79/77-78, 80/79/77-78, 121-122, 123/121-122, 135-136, 137/135-136, 138/137/135-136
Y	US 2016/0153005 A1 (THE BROAD INSTITUTE INC., et al.) 02 June 2016; abstract; figure 79; paragraphs [0197], [0243], [0247], [0543], [1187], [1373]	77-78, 79/77-78, 80/79/77-78, 121-122, 123/121-122, 135-136, 137/135-136, 138/137/135-136

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
19 June 2020 (19.06.2020)	16 JUL 2020
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Shane Thomas Telephone No. PCT Helpdesk: 571-272-4300

Form PCT/ISA/210 (second sheet) (July 2019)

RECEIVED
Wolf, Greenfield & Sacks, P.C.
JUL 20 2020
Docketed ____ Already Docketed ____
Not Required ____
Initials 1st ____ 2nd ____

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To: Tani Chen
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206

PCT

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing *(day/month/year)*	16 JUL 2020

Applicant's or agent's file reference	FOR FURTHER ACTION
H0498.70693WO00	See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US20/28632	17 April 2020 (17.04.2020)	19 April 2019 (19.04.2019)

International Patent Classification (IPC) or both national classification and IPC
IPC - C12Q 1/6832, 1/68; C12N 9/22, 15/11 (2020.01)
CPC - C12N 9/22, 15/11; C12Q 1/6832, 1/6874

Applicant
PRESIDENT AND FELLOWS OF HARVARD COLLEGE

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☒ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☐ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☐ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA/US	Date of completion of this opinion	Authorized officer
Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	19 June 2020 (19.06.2020)	Shane Thomas PCT Help Desk Telephone No. 571-272-4300

Form PCT/ISA/237 (cover sheet) (revised January 2019)

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US20/28632

Box No. I Basis of this opinion

1. With regard to the **language**, this opinion has been established on the basis of:

☒ the international application in the language in which it was filed.

☐ a translation of the international application into _____, which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(b)).

3. ☒ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:

a. ☒ forming part of the international application as filed:

☒ in the form of an Annex C/ST.25 text file.

☐ on paper or in the form of an image file.

b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US20/28632

Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☒ claims Nos. 4-75, 81-120, 124-134, 139-179

because:

☐ the said international application, or the said claims Nos. _____ relate to the following subject matter which does not require an international search *(specify)*:

☒ the description, claims or drawings *(indicate particular elements below)* or said claims Nos. -***-Please See Below-***- are so unclear that no meaningful opinion could be formed *(specify)*.

Claims Nos.: 4-75, 81-120, 124-134, 139-179 are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

☐ the claims, or said claims Nos. _____ are so inadequately supported by the description that no meaningful opinion could be formed *(specify)*:

☒ no international search report has been established for said claims Nos. 4-75, 81-120, 124-134, 139-179

☐ a meaningful opinion could not be formed without the sequence listing; the applicant did not, within the prescribed time limit:

☐ furnish a sequence listing in the form of an Annex C/ST.25 text file, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.

☐ furnish a sequence listing on paper or in the form of an image file complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.

☐ pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rule 13*ter*.1(a) or (b).

☐ See Supplemental Box for further details.

Form PCT/ISA/237 (Box No. III) (revised January 2019)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US20/28632

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	-***-Please See Below-***-	YES
	Claims	NONE	NO
Inventive step (IS)	Claims	NONE	YES
	Claims	-***-Please See Below-***-	NO
Industrial applicability (IA)	Claims	-***-Please See Below-***-	YES
	Claims	NONE	NO

2. Citations and explanations.

-Continued from Box No. V: 1. Statement:

Novelty (N): YES: Claims 1-2, 3/1-2, 76-78, 79/77-78, 80/79/77-78, 121-122, 123/121-122, 135-136, 137/135-136, 138/137/135-136

Inventive step (IS): NO: Claims 1-2, 3/1-2, 76-78, 79/77-78, 80/79/77-78, 121-122, 123/121-122, 135-136, 137/135-136, 138/137/135-136

Industrial Applicability (IA): YES: Claims 1-2, 3/1-2, 76-78, 79/77-78, 80/79/77-78, 121-122, 123/121-122, 135-136, 137/135-136, 138/137/135-136

-Continued from Box No. V: 2. Citations and Explanations:

Claims 1-2, 3/1-2, and 76 lack an inventive step under PCT Article 33(3) as being obvious over the publication, 'Spatially Resolved, Highly Multiplexed RNA Profiling in Single Cells' to Chen et al. (hereinafter 'Chen') in view of US 2018/0320226 A1 (President and Fellows of Harvard College') (hereinafter 'Harvard').

As per claim 1, Chen discloses a method (abstract; page 412, column 2, second paragraph), comprising: (a) introducing, into a plurality of cells, a recognition sequence, a reporter portion, and an identification portion comprising read sequences (we labeled each cellular RNA in human fibroblast cells with a set of encoding probes, which contain targeting sequences that bind the RNA (a recognition sequence) and readout sequences that bind fluorescently labeled readout probes (a reporter portion), wherein each RNA species is encoded with a particular combination of readout sequences (an identification portion comprising read sequences); page 412, column 2, second paragraph; column 3, second paragraph; figure 1; figure 1 legend); (b) determining positions of RNA molecules expressed from the reporter portion by determining the reporter portions (determining positions of RNA molecules expressed from the reporter portion by determining the reporter portions; page 412, column 2, second paragraph; figure 1; figure 1 legend); (c) determining a read sequence on the RNA molecules expressed from the introduced DNA comprising the reporter portion and the identification portion within the plurality of cells by exposing the cells to a readout probe able to bind to the read sequence (determining a read sequence on the RNA molecules expressed from the introduced DNA comprising the reporter portion and the identification portion within the plurality of cells by exposing the cells to a readout probe able to bind to the read sequence; page 412, column 2, second paragraph; figure 1; figure 1 legend); (d) colocalizing the binding of the readout probe with the positions of the RNA molecules expressed from the reporter portion of the introduced DNA (combinatorial labeling allows the number of detectable RNA species using hybridization and imaging, each with a different readout probe; page 412, column 2, second paragraph; figure 1; figure 1 legend); (e) repeating (b), (c), and (d) a plurality of times using different read sequences (performing successive rounds; page 412, column 2, second paragraph; page 412, column 2, second paragraph figure 1; figure 1 legend); and (f) creating codewords corresponding to the binding of the colocalized readout probes, wherein the values of the digits of the codewords are based on the binding of the readout probes to the read sequences (we then constructed binary code words from the observed fluorescent spots based on their on-off patterns across the 16 hybridization rounds; page aaa6090-2, column 2, third paragraph; figure 1; figure 1 legend). Chen does not disclose DNA comprising a guide portion. Harvard discloses a method (abstract), comprising: (a) introducing, into a plurality of cells (introducing, into a plurality of cells; paragraphs [0072], [0187]), DNA comprising a guide portion (a complex including a guide RNA, a DNA binding protein, such as a Cas9 protein; paragraph [0004]) comprising a recognition sequence (a guide portion comprising a recognition sequence; paragraphs [0169], [0193]), a reporter portion (guide RNA includes a 3' tail sequence comprising one or more detectable labels, such as a fluorescent nucleotide; paragraphs [0008], [0190]), and an identification portion comprising read sequences (and an barcode; paragraph [0190], [0191]); (b) determining positions of RNA molecules expressed from the reporter portion of the introduced DNA within the plurality of cells by determining the reporter portions (the Cas9 and guide RNA are complexed before being used to target the DNA, the guide RNA is complementary to the target DNA, and the complex bound to DNA is detected by measuring fluorescent signal; paragraph [0027]); (c) determining a read sequence on the RNA molecules expressed from the introduced DNA comprising the reporter portion and the identification portion within the plurality of cells by exposing the cells to a readout probe able to bind to the read sequence (by decoding each barcode, we identify each gRNA/Cas9 and its location on the locus, which provides information about insertions, deletions and regions rearranged within its locus; paragraph [0234]). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Chen, to include DNA comprising a guide portion, as previously disclosed by Harvard, in order to provide a method that allows for multiple specific locations to be targeted with the guide RNA (Harvard; paragraph [0015]).

-***-Continued Within the Next Supplemental Box-***-

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US20/28632

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of.

-***-Continued from Box V: Citations and Explanations-***-

As per claim 2, Chen and Harvard, in combination, disclose the method of claim 1, and Chen further discloses identifying the guide portion for individual cells based on the measured codewords (if the word exactly matched one of the 140 MHD4 code words (exact matches) or differed by only one bit (error-correctable matches), we assigned it to the corresponding RNA species; page aaa6090-2, column 2, third paragraph; figure 1; figure 1 legend). Chen does not disclose comprising a guide portion. Harvard discloses a guide portion (a complex including a guide RNA, a DNA binding protein, such as a Cas9 protein; paragraph [0004]). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Chen, to include a guide portion, as previously disclosed by Harvard, in order to provide a method that allows for multiple specific locations to be targeted with the guide RNA (Harvard; paragraph [0015]).

As per claims 3/1-2, Chen and Harvard, in combination, disclose the method of claims 1-2, but Chen does not disclose wherein the recognition sequence recognizes a DNA sequence. Harvard discloses wherein the recognition sequence recognizes a DNA sequence (the Cas9 and guide RNA are complexed before being used to target the DNA, the guide RNA is complementary to the target DNA, and the complex bound to DNA is detected by measuring fluorescent signal; paragraph [0027]). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Chen, to include wherein the recognition sequence recognizes a DNA sequence, as previously disclosed by Harvard, in order to provide a method that allows for multiple specific locations to be targeted with the guide RNA (Harvard; paragraph [0015]), such as DNA.

As per claim 76, Chen discloses a method (abstract; page 412, column 2, second paragraph), comprising: introducing, into a plurality of cells, a recognition sequence, a reporter portion, and an identification portion comprising read sequences (we labeled each cellular RNA in human fibroblast cells with a set of encoding probes, which contain targeting sequences that bind the RNA (a recognition sequence) and readout sequences that bind fluorescently labeled readout probes (a reporter portion), wherein each RNA species is encoded with a particular combination of readout sequences (an identification portion comprising read sequences); page 412, column 2, second paragraph; column 3, second paragraph; figure 1; figure 1 legend); determining positions of RNA molecules expressed from the reporter portion the plurality of cells by determining the reporter portions (determining positions of RNA molecules expressed from the reporter portion by determining the reporter portions; page 412, column 2, second paragraph; figure 1; figure 1 legend); determining the read sequences within the plurality of cells by exposing the cells to a plurality of readout probes each able to bind to a read sequence (determining a read sequence on the RNA molecules expressed from the introduced DNA comprising the reporter portion and the identification portion within the plurality of cells by exposing the cells to a readout probe able to bind to the read sequence; page 412, column 2, second paragraph; figure 1; figure 1 legend), colocalizing the binding of the readout probes with the positions of the RNA molecules expressed from the reporter portion of the introduced DNA (combinatorial labeling allows the number of detectable RNA species using hybridization and imaging, each with a different readout probe; page 412, column 2, second paragraph; figure 1; figure 1 legend); and creating codewords corresponding to the binding of the colocalized readout probes, wherein the values of the digits of the codewords are based on the binding of the readout probes to the read sequences (we then constructed binary code words from the observed fluorescent spots based on their on-off patterns across the 16 hybridization rounds; page aaa6090-2, column 2, third paragraph; figure 1; figure 1 legend). Chen does not disclose DNA comprising a guide portion. Harvard discloses a method (abstract), comprising: introducing, into a plurality of cells (introducing, into a plurality of cells; paragraphs [0072], [0167], DNA comprising a guide portion (a complex including a guide RNA, a DNA binding protein, such as a Cas9 protein; paragraph [0004]) comprising a recognition sequence (a guide portion comprising a recognition sequence; paragraphs [0169], [0193]), a reporter portion (guide RNA includes a 3′ tail sequence comprising one or more detectable labels, such as a fluorescent nucleotide; paragraphs [0009], [0190]), and an identification portion comprising read sequences (and an barcode; paragraph [0190], [0191]); determining positions of RNA molecules expressed from the reporter portion of the introduced DNA within the plurality of cells by determining the reporter portions (the Cas9 and guide RNA are complexed before being used to target the DNA, the guide RNA is complementary to the target DNA, and the complex bound to DNA is detected by measuring fluorescent signal; paragraph [0027]); determining the read sequences within the plurality of cells by exposing the cells to a plurality of readout probes each able to bind to a read sequence (by decoding each barcode, we identify each gRNA/Cas9 and its location on the locus, which provides information about insertions, deletions and regions rearranged within its locus; paragraph [0234]). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Chen, to include DNA comprising a guide portion, as previously disclosed by Harvard, in order to provide a method that allows for multiple specific locations to be targeted with the guide RNA (Harvard; paragraph [0015]).

-***-Continued Within the Next Supplemental Box-***-

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US20/28632

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

-***-Continued from Previous Supplemental Box-***-

Claims 77-78, 79/77-78, 80/79/77-78, 121-122, 123/121-122, 135-136, 137/135-136, and 138/137/135-136 lack an inventive step under PCT Article 33(3) as being obvious over Harvard in view of US 2016/0153005 A1 to The Broad Institute Inc. et al. (hereinafter 'Broad').

As per claim 77, Harvard discloses a method (abstract), comprising: introducing DNA into a plurality of cells (introducing, into a plurality of cells using a vector; paragraphs [0072], [0167], wherein the DNA comprises a guide portion (a complex including a guide RNA, a DNA binding protein, such as a Cas9 protein; paragraph [0004]) comprising a recognition sequence (a guide portion comprising a recognition sequence; paragraphs [0169], [0193]), a reporter portion (guide RNA includes a 3′ tail sequence comprising one or more detectable labels, such as a fluorescent nucleotide; paragraphs [0009], [0190]), and an identification portion comprising read sequences (guide RNA includes a 3′ tail sequence comprising one or more detectable labels, such as a fluorescent nucleotide; paragraphs [0009], [0190]); and determining genotype of the plurality of cells (whereby the sequence of the target nucleic acid is determinad by sequencing methods; paragraph [0015]). Harvard does not disclose using a lentivirus; determining phenotype of the plurality of cells; and determining the correspondence between the genotype and the phenotype. Broad discloses using a lentivirus (paragraph [0048]); determining phenotype of the plurality of cells (paragraphs [0243], [0247]); and determining the correspondence between the genotype and the phenotype (not only was a genotypic change seen in a "gold-standard" gene for liver such as ApoB, but phenotypic changes were also recorded; paragraphs [0243], [0247]). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Harvard, to include using a lentivirus; determining phenotype of the plurality of cells; and determining the correspondence between the genotype and the phenotype, as previously disclosed by Broad, in order to ensure enhanced specificity for target recognition and avoidance of toxicity and to edit or modify a target site in a genomic locus of interest to alter or improve the status of a disease or a condition (Broad; abstract).

As per claim 78, Harvard and Broad, in combination, disclose the method of claim 77, but Harvard does not disclose wherein determining phenotype comprises determining a cell property. Broad discloses wherein determining phenotype comprises determining a cell property (applicants tested overall morphology of neurons; paragraph [1187]). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Harvard, to include wherein determining phenotype comprises determining a cell, as previously disclosed by Broad, in order to ensure enhanced specificity for target recognition and avoidance of toxicity and to edit or modify a target site in a genomic locus of interest to alter or improve the status of a disease or a condition (Broad; abstract).

As per claims 79/77-78, Harvard and Broad, in combination, disclose the method of claims 77-78, but Harvard does not disclose wherein determining phenotype comprises determining morphology of a cellular structure. Broad discloses wherein determining phenotype comprises determining morphology of a cellular structure (applicants tested overall morphology of neurons; paragraph [1187]). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Harvard, to include wherein determining phenotype comprises determining morphology of a cellular structure, as previously disclosed by Broad, in order to ensure enhanced specificity for target recognition and avoidance of toxicity and to edit or modify a target site in a genomic locus of interest to alter or improve the status of a disease or a condition (Broad; abstract).

As per claims 80/79/77-78, Harvard and Broad, in combination, disclose the method of claims 79/77-78, but Harvard does not disclose wherein the morphology is whole cell morphology. Broad discloses wherein the morphology is whole cell morphology (applicants tested overall morphology of neurons; paragraph [1187]). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Harvard, to include wherein the morphology is whole cell morphology, as previously disclosed by Broad, in order to ensure enhanced specificity for target recognition and avoidance of toxicity and to edit or modify a target site in a genomic locus of interest to alter or improve the status of a disease or a condition (Broad; abstract).

As per claim 121, Harvard discloses a method (abstract), comprising: introducing nucleic acids into a plurality of cells (introducing, into a plurality of cells using a vector; paragraphs [0072], [0167]), wherein the nucleic acids comprise a guide portion (a complex including a guide RNA, a DNA binding protein, such as a Cas9 protein; paragraph [0004]) comprising a recognition sequence (a guide portion comprising a recognition sequence; paragraphs [0169], [0193]), a reporter portion (guide RNA includes a 3′ tail sequence comprising one or more detectable labels, such as a fluorescent nucleotide; paragraphs [0009], [0190]), and an identification portion comprising read sequences (guide RNA includes a 3′ tail sequence comprising one or more detectable labels, such as a fluorescent nucleotide; paragraphs [0009], [0190]); imaging the plurality of cells (FISH imaging; paragraph [0070]). Harvard does not disclose wherein the cells exhibit imagable differences in phenotype due to expression of the guide portion; and acquiring a plurality of images of the plurality of cells, wherein the images of the cells exhibit differences due to differences in the identification portions of the nucleic acids within the cells. Broad discloses wherein the cells exhibit imagable differences in phenotype due to expression of the guide portion (CRISPR-Cas9-mediated genome editing can be used to correct a disease mutation and/or phenotype, wherein phenotypic changes were seen with two of the three guides used (targets 1 and 3), as seen in FIG. 57B, which shows oil red staining to detect hepatic lipid accumulation phenotype in vivo following AAV-Cas9-sgRNA delivery; paragraphs [0543], [1373]); and acquiring a plurality of images of the plurality of cells, wherein the images of the cells exhibit differences due to differences in the identification portions of the nucleic acids within the cells (and acquiring a plurality of images of the plurality of cells, wherein the images of the cells exhibit differences due to differences in the identification portions of the nucleic acids within the cells; figure 79; paragraphs [0197], [1373]). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Harvard, to include wherein the cells exhibit imagable differences in phenotype due to expression of the guide portion; and acquiring a plurality of images of the plurality of cells, wherein the images of the cells exhibit differences due to differences in the identification portions of the nucleic acids within the cells, as previously disclosed by Broad, in order to ensure enhanced specificity for target recognition and avoidance of toxicity and to edit or modify a target site in a genomic locus of interest to alter or improve the status of a disease or a condition (Broad; abstract).

-***-Continued Within the Next Supplemental Box-***-

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US20/28632

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**
Continuation of:

-***-Continued from Previous Supplemental Box-***-

As per claim 122, Harvard and Broad, in combination, disclose the method of claim 121, but Harvard does not disclose wherein based on the correspondence between the guide portion and the identification portion, the guide portion introduced to the plurality of cells is identified. Broad discloses wherein based on the correspondence between the guide portion and the identification portion, the guide portion introduced to the plurality of cells is identified (and acquiring a plurality of images of the plurality of cells, wherein the images of the cells exhibit differences due to differences in the identification portions of the nucleic acids within the cells; figure 79; paragraphs [0197], [1373]). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Harvard, to include wherein the cells exhibit imagable differences in phenotype due to expression of the guide portion; and acquiring a plurality of images of the plurality of cells, wherein the images of the cells exhibit differences due to differences in the identification portions of the nucleic acids within the cells, as previously disclosed by Broad, in order to ensure enhanced specificity for target recognition and avoidance of toxicity and to edit or modify a target site in a genomic locus of interest to alter or improve the status of a disease or a condition (Broad; abstract).

As per claim 123/121-122, Harvard and Broad, in combination, disclose the method of claims 121-122, but Harvard does not disclose wherein cells having different phenotypes have different appearances when imaged. Broad discloses wherein cells having different phenotypes have different appearances when imaged (wherein cells having different phenotypes have different appearances when imaged; figure 79; paragraphs [0197], [1373]). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Harvard, to include wherein the cells exhibit imagable differences in phenotype due to expression of the guide portion; and acquiring a plurality of images of the plurality of cells, wherein the images of the cells exhibit differences due to differences in the identification portions of the nucleic acids within the cells, as previously disclosed by Broad, in order to ensure enhanced specificity for target recognition and avoidance of toxicity and to edit or modify a target site in a genomic locus of interest to alter or improve the status of a disease or a condition (Broad; abstract).

As per claim 135, Harvard discloses a method (abstract), comprising: introducing DNA into a plurality of cells (introducing, into a plurality of cells using a vector; paragraphs [0072], [0167]), wherein the DNA comprises a guide portion (a complex including a guide RNA, a DNA binding protein, such as a Cas9 protein; paragraph [0004]) comprising a recognition sequence (a guide portion comprising a recognition sequence; paragraphs [0169], [0193]) and an identification portion comprising read sequences (guide RNA includes a 3′ tail sequence comprising one or more detectable labels, such as a fluorescent nucleotide; paragraphs [0009], [0190]); determining genotype of the plurality of cells (whereby the sequence of the target nucleic acid is determined by sequencing methods; paragraph [0015]). Harvard does not disclose using a lentivirus; determining phenotype of the plurality of cells; and determining the correspondence between genotypes and phenotypes. Broad discloses using a lentivirus (paragraph [0048]); determining phenotype of the plurality of cells (paragraphs [0243], [0247]); and determining the correspondence between genotypes and phenotypes (not only was a genotypic change seen in a "gold-standard" gene for liver such as ApoB, but phenotypic changes were also recorded; paragraphs [0243], [0247]). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Harvard, to include using a lentivirus; determining phenotype of the plurality of cells; and determining the correspondence between the genotype and the phenotype, as previously disclosed by Broad, in order to ensure enhanced specificity for target recognition and avoidance of toxicity and to edit or modify a target site in a genomic locus of interest to alter or improve the status of a disease or a condition (Broad; abstract).

As per claim 136, Harvard and Broad, in combination, disclose the method of claim 135, but Harvard does not disclose wherein determining phenotype comprises determining a cell property. Broad discloses wherein determining phenotype comprises determining a cell property (applicants tested overall morphology of neurons; paragraph [1187]). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Harvard, to include wherein determining phenotype comprises determining a cell, as previously disclosed by Broad, in order to ensure enhanced specificity for target recognition and avoidance of toxicity and to edit or modify a target site in a genomic locus of interest to alter or improve the status of a disease or a condition (Broad; abstract).

As per claims 137/135-136, Harvard and Broad, in combination, disclose the method of claims 135-136, but Harvard does not disclose wherein determining phenotype comprises determining morphology of a cellular structure. Broad discloses wherein determining phenotype comprises determining morphology of a cellular structure (applicants tested overall morphology of neurons; paragraph [1187]). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Harvard, to include wherein determining phenotype comprises determining morphology of a cellular structure, as previously disclosed by Broad, in order to ensure enhanced specificity for target recognition and avoidance of toxicity and to edit or modify a target site in a genomic locus of interest to alter or improve the status of a disease or a condition (Broad; abstract).

As per claims 138/137/135-136, Harvard and Broad, in combination, disclose the method of claims 137/135-136, but Harvard does not disclose wherein the morphology is whole cell morphology. Broad discloses wherein the morphology is whole cell morphology (applicants tested overall morphology of neurons; paragraph [1187]). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Harvard, to include wherein the morphology is whole cell morphology, as previously disclosed by Broad, in order to ensure enhanced specificity for target recognition and avoidance of toxicity and to edit or modify a target site in a genomic locus of interest to alter or improve the status of a disease or a condition (Broad; abstract).

Claims 1-2, 3/1-2, 76-78, 79/77-78, 80/79/77-78, 121-122, 123/121-122, 135-136, 137/135-136, and 138/137/135-136 have industrial applicability as defined by PCT Article 33(4) because the subject matter can be made or used in industry.

PCT Recordation of Search History

Case/PCT Application Number: PCT/US20/28632

CLIN Number/Technical Field of PCT Application: 0007 Biotechnology

Date(s) During Which the Search was Conducted: 19 June 2020

Date of Completion of Recordation of Search History Form: 19 June 2020

Research Analyst Initials: CJF

Search Approval Official (SAO) Initials: DGF

Field of Search/Classification Information:

IPC(8): C12Q 1/6832, 1/68; C12N 9/22, 15/11 (2020.01)

CPC: C12N 9/22, 15/11, 2310/20, 2800/80; C12Q 1/6832, 1/6874, 2521/301, 2543/10, 2563/107, 2565/601

Databases Searched (Patent and Non-Patent Literature (NPL), Including Sub-Databases and Files Searched) and Search Terms Used:

PatSeer (US, EP, WO, JP, DE, GB, CN, FR, KR, ES, AU, IN, CA, INPADOC Data); EBSCO Discovery; PubMed; Microsoft Academic; Patseer;

KEYWORDS: guide sequence, recognition sequence, reporter portion, identification portion, read sequence, genotype, phenotype

Database Search String Recordation, Including Dates of Searches:

Patent Database Search Strategy/Results:

PatSeer

ID	Search String	Records	Created Date
L27	TACD: (guide rna* OR sgrna* OR grna*) AND recog* AND (reporter* OR fluor*) AND identif* AND read* AND sequ* AND ((ASN: "PRESIDENT AND FELLOWS OF HARVARD COLLEGE") OR (INV: ZHUANG Xiaowei) OR (INV: WANG Chong))	57	19-Jun-20
L26	TAC: (guide rna* OR sgrna* OR grna*) AND recog* AND (reporter* OR fluor*) AND identif* AND read* AND sequ* AND ((ASN: "PRESIDENT AND FELLOWS OF HARVARD COLLEGE") OR (INV: ZHUANG Xiaowei) OR (INV: WANG Chong))	0	19-Jun-20
L25	TA: (guide rna* OR sgrna* OR grna*) AND recog* AND (reporter* OR fluor*) AND identif* AND read* AND sequ* AND ((ASN: "PRESIDENT AND FELLOWS OF HARVARD COLLEGE") OR (INV: ZHUANG Xiaowei) OR (INV: WANG Chong))	0	19-Jun-20
L24	TACD: (guide rna* OR sgrna* OR grna*) AND recog* AND (reporter* OR fluor*) AND identif* AND read* AND sequ* AND genotyp* AND phenotyp* AND ((IC: C12Q1/6832 OR C12Q1/68 OR C12N9/22 OR C12N15/11) OR (CPC: C12N9/22 OR C12N15/11 OR C12N2310/20 OR C12N2800/80 OR C12Q1/6832 OR C12Q1/6874 OR C12Q2521/301 OR C12Q2543/10 OR C12Q2563/107 OR C12Q2565/601))	3765	19-Jun-20
L23	TACD: (guide rna* OR sgrna* OR grna*) AND recog* AND (reporter* OR fluor*) AND identif* AND read* AND sequ* AND genotyp* AND phenotyp*	7795	19-Jun-20
L22	TAC: (guide rna* OR sgrna* OR grna*) AND recog* AND (reporter* OR fluor*) AND identif* AND read* AND sequ* AND genotyp* AND phenotyp*	4	19-Jun-20
L21	TA: (guide rna* OR sgrna* OR grna*) AND recog* AND (reporter* OR fluor*) AND identif* AND read* AND sequ* AND genotyp* AND phenotyp*	0	19-Jun-20
L20	TACD: (guide rna* OR sgrna* OR grna*) AND recog* AND (reporter* OR fluor*) AND identif* AND read* AND sequ* and (code* AND word OR codeword*) AND ((IC: C12Q1/6832 OR C12Q1/68 OR C12N9/22 OR C12N15/11) OR (CPC: C12N9/22 OR C12N15/11 OR C12N2310/20 OR C12N2800/80 OR C12Q1/6832 OR C12Q1/6874 OR C12Q2521/301 OR C12Q2543/10 OR C12Q2563/107 OR C12Q2565/601))	6372	19-Jun-20
L19	TACD: (guide rna* OR sgrna* OR grna*) AND recog* AND (reporter* OR fluor*) AND identif* AND read* AND sequ* and (code* AND word OR codeword*)	13495	19-Jun-20
L18	TAC: (guide rna* OR sgrna* OR grna*) AND recog* AND (reporter* OR fluor*) AND identif* AND read* AND sequ* and (code* AND word OR codeword*)	7	19-Jun-20
L17	TA: (guide rna* OR sgrna* OR grna*) AND recog* AND (reporter* OR fluor*) AND identif* AND read* AND sequ* and (code* AND word OR codeword*)	0	19-Jun-20
L16	TACD: (guide rna* OR sgrna* OR grna*) WP recog* WP (reporter* OR fluor*) WP identif* WP read* WP sequ* and (code* WP word OR codeword*)	0	19-Jun-20

L15	TAC: (guide rna* OR sgrna* OR grna*) WP recog* WP (reporter* OR fluor*) WP identif* WP read* WP sequ* and (code* WP word OR codeword*)	0	19-Jun-20
L14	TA: (guide rna* OR sgrna* OR grna*) WP recog* WP (reporter* OR fluor*) WP identif* WP read* WP sequ* and (code* WP word OR codeword*)	0	19-Jun-20
L13	TACD: (guide rna* OR sgrna* OR grna*) W50 recog* W50 (reporter* OR fluor*) W50 identif* W50 read* W50 sequ* and (code* W50 word OR codeword*)	0	19-Jun-20
L12	TAC: (guide rna* OR sgrna* OR grna*) W50 recog* W50 (reporter* OR fluor*) W50 identif* W50 read* W50 sequ* and (code* W50 word OR codeword*)	0	19-Jun-20
L11	TA: (guide rna* OR sgrna* OR grna*) W50 recog* W50 (reporter* OR fluor*) W50 identif* W50 read* W50 sequ* and (code* W50 word OR codeword*)	0	19-Jun-20
L10	TACD: (guide rna* OR sgrna* OR grna*) W20 recog* W20 (reporter* OR fluor*) W20 identif* W20 read* W20 sequ* and (code* W20 word OR codeword*)	0	19-Jun-20
L9	TAC: (guide rna* OR sgrna* OR grna*) W20 recog* W20 (reporter* OR fluor*) W20 identif* W20 read* W20 sequ* and (code* W20 word OR codeword*)	0	19-Jun-20
L8	TA: (guide rna* OR sgrna* OR grna*) W20 recog* W20 (reporter* OR fluor*) W20 identif* W20 read* W20 sequ* and (code* W20 word OR codeword*)	0	19-Jun-20
L7	TACD: (guide rna* OR sgrna* OR grna*) Wp recog* Wp (reporter* OR fluor*) Wp identif* Wp read* Wp sequ*	29	19-Jun-20
L6	TACD: (guide rna* OR sgrna* OR grna*) W50 recog* W50 (reporter* OR fluor*) W50 identif* W50 read* W50 sequ*	0	19-Jun-20
L5	TAC: (guide rna* OR sgrna* OR grna*) W50 recog* W50 (reporter* OR fluor*) W50 identif* W50 read* W50 sequ*	0	19-Jun-20
L4	TA: (guide rna* OR sgrna* OR grna*) W50 recog* W50 (reporter* OR fluor*) W50 identif* W50 read* W50 sequ*	0	19-Jun-20
L3	TACD: (guide rna* OR sgrna* OR grna*) W20 recog* W20 (reporter* OR fluor*) W20 identif* W20 read* W20 sequ*	0	19-Jun-20
L2	TAC: (guide rna* OR sgrna* OR grna*) W20 recog* W20 (reporter* OR fluor*) W20 identif* W20 read* W20 sequ*	0	19-Jun-20
L1	TA: (guide rna* OR sgrna* OR grna*) W20 recog* W20 (reporter* OR fluor*) W20 identif* W20 read* W20 sequ*	0	19-Jun-20

Non-Patent Literature (NPL) Search Strategy/Results:

Microsoft Academic: 19 June 2020

Search String	Results
guide sequence recognition sequence reporter portion identification portion read sequence fluorescent code word	491
introduce DNA guide portion recognition sequence reporter portion identification portion	184
guide sequence recognition sequence reporter portion identification portion read sequence genotype phenotype	18

guide sequence recognition sequence reporter portion (ZHUANG, Xiaowei) (WANG, Chong)	47

PubMed: 19 June 2020

Search String	Results
guide sequence recognition sequence reporter portion identification portion read sequence fluorescent code word	0
introduce DNA guide portion recognition sequence reporter portion identification portion	0
guide sequence recognition sequence reporter portion identification portion read sequence genotype phenotype	0

EBSCO Discovery: 19 June 2020

Search String	Results
guide sequence recognition sequence reporter portion identification portion read sequence fluorescent code word	19907
introduce DNA guide portion recognition sequence reporter portion identification portion	41945
guide sequence recognition sequence reporter portion identification portion read sequence genotype phenotype	14489

Wolf, Greenfield & Sacks, P.C.

PATENT COOPERATION TREATY

FD MAY 04 2020
Docketed ✓ Already Docketed
Not Required
Initials To 2nd

From the INTERNATIONAL SEARCHING AUTHORITY

To: Tani Chen
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
United States of America

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT AND THE WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY, OR THE DECLARATION

(PCT Rule 44.1)

	Date of mailing *(day/month/year)*: 01 MAY 2020
Applicant's or agent's file reference: H0498.70680WO00	**FOR FURTHER ACTION** See paragraphs 1 and 4 below
International application No.: PCT/US19/65857	International filing date *(day/month/year)*: 12 December 2019 (12.12.2019)
Applicant: PRESIDENT AND FELLOWS OF HARVARD COLLEGE	

1. ☒ The applicant is hereby notified that the international search report and the written opinion of the International Searching Authority have been established and are transmitted herewith.

   Filing of amendments and statement under Article 19:
   The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46):

   When? The time limit for filing such amendments is normally two months from the date of transmittal of the international search report.

   How? Directly to the International Bureau preferably through ePCT, or on paper to:
   The International Bureau of WIPO, 34, chemin des Colombettes, 1211 Geneva 20, Switzerland

   For more detailed instructions, see the *PCT Applicant's Guide*, International Phase, paragraphs 9.004–9.011.

2. ☐ The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect and the written opinion of the International Searching Authority are transmitted herewith.

3. ☐ **With regard to any protest** against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:
   ☐ the protest together with the decision thereon has been transmitted to the International Bureau together with any request to forward the texts of both the protest and the decision thereon to the designated Offices.
   ☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Reminders**

   The applicant may **submit comments on an informal basis on the written opinion of the International Searching Authority** to the International Bureau. These comments will be made available to the public after international publication. The International Bureau will send a copy of such comments to all designated Offices unless an international preliminary examination report has been or is to be established.

   Shortly after the expiration of **18 months from the priority date, the international application will be published** by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau before the completion of the technical preparations for international publication (Rules 90*bis*.1 and 90*bis*.3).

   Within **19 months** from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase **until 30 months** from the priority date (in some Offices even later); otherwise, the applicant must, **within 20 months** from the priority date, perform the prescribed acts for **entry into the national phase** before those designated Offices. In respect of other designated Offices, the time limit of **30 months** (or later) will apply even if no demand is filed within 19 months. For details about the applicable time limits, Office by Office, see www.wipo.int/pct/en/texts/time_limits.html and the *PCT Applicant's Guide*, National Chapters.

   Within **22 months from the priority date, the applicant may request that a supplementary international search be carried out** by a different International Searching Authority that offers this service (Rule 45*bis*.1). The procedure for requesting supplementary international search is described in the *PCT Applicant's Guide*, International Phase, paragraphs 8.006-8.032.

Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Shane Thomas Telephone No. PCT Helpdesk: 571-272-4300

Form PCT/ISA/220 (revised January 2020)

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

To: Tani Chen
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
United States of America

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT AND THE WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY, OR THE DECLARATION

(PCT Rule 44.1)

Applicant's or agent's file reference H0498.70680WO00	Date of mailing *(day/month/year)* **FOR FURTHER ACTION** See paragraphs 1 and 4 below
International application No. PCT/US19/65857	International filing date *(day/month/year)* 12 December 2019 (12.12.2019)
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE	

1. ☒ The applicant is hereby notified that the international search report and the written opinion of the International Searching Authority have been established and are transmitted herewith.

   Filing of amendments and statement under Article 19:
   The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46):

   When? The time limit for filing such amendments is normally two months from the date of transmittal of the international search report.

   How? Directly to the International Bureau preferably through ePCT, or on paper to:
   The International Bureau of WIPO, 34, chemin des Colombettes, 1211 Geneva 20, Switzerland

   For more detailed instructions, see the *PCT Applicant's Guide*, International Phase, paragraphs 9.004–9.011.

2. ☐ The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect and the written opinion of the International Searching Authority are transmitted herewith.

3. ☐ **With regard to any protest** against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:
   - ☐ the protest together with the decision thereon has been transmitted to the International Bureau together with any request to forward the texts of both the protest and the decision thereon to the designated Offices.
   - ☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Reminders**

   The applicant may **submit comments on an informal basis on the written opinion of the International Searching Authority** to the International Bureau. These comments will be made available to the public after international publication. The International Bureau will send a copy of such comments to all designated Offices unless an international preliminary examination report has been or is to be established.

   Shortly after the expiration of **18 months from the priority date, the international application will be published** by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau before the completion of the technical preparations for international publication (Rules 90*bis*.1 and 90*bis*.3).

   Within **19 months** from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase **until 30 months** from the priority date (in some Offices even later); otherwise, the applicant must, **within 20 months** from the priority date, perform the prescribed acts for **entry into the national phase** before those designated Offices. In respect of other designated Offices, the time limit of **30 months** (or later) will apply even if no demand is filed within 19 months. For details about the applicable time limits, Office by Office, see www.wipo.int/pct/en/texts/time_limits.html and the *PCT Applicant's Guide*, National Chapters.

   Within **22 months from the priority date, the applicant may request that a supplementary international search be carried out** by a different International Searching Authority that offers this service (Rule 45*bis*.1). The procedure for requesting supplementary international search is described in the *PCT Applicant's Guide*, International Phase, paragraphs 8.006-8.032.

Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Shane Thomas Telephone No. PCT Helpdesk: 571-272-4300

Form PCT/ISA/220 (revised January 2020)

RECEIVED
Wolf, Greenfield & Sacks, PC
MAY 04 2020
Docketed ____ Already Docketed ____
Not Required ____
Initials 1st ____ 2nd ____

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference	FOR FURTHER ACTION
H0498.70660WO00	see Form PCT/ISA/220 as well as, where applicable, item 5 below.

International application No.	International filing date *(day/month/year)*	(Earliest) Priority Date *(day/month/year)*
PCT/US19/65857	12 December 2019 (12.12.2019)	13 December 2018 (13.12.2018)

Applicant
PRESIDENT AND FELLOWS OF HARVARD COLLEGE

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 6 sheets.

☐ It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

a. With regard to the **language**, the international search was carried out on the basis of:

☒ the international application in the language in which it was filed.

☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

b. ☐ This international search report has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43.6*bis*(a)).

c. ☒ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, see Box No. I.

2. ☒ **Certain claims were found unsearchable** (see Box No. II).

3. ☒ **Unity of invention is lacking** (see Box No. III).

4. With regard to the **title**,

☒ the text is approved as submitted by the applicant.

☐ the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

☒ the text is approved as submitted by the applicant.

☐ the text has been established, according to Rule 38.2, by this Authority as it appears in Box No. IV. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. With regard to the **drawings**,

a. the figure of the drawings to be published with the abstract is Figure No. 1D

☒ as suggested by the applicant.

☐ as selected by this Authority, because the applicant failed to suggest a figure.

☐ as selected by this Authority, because this figure better characterizes the invention.

b. ☐ none of the figures is to be published with the abstract.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US19/65857

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. ☒ forming part of the international application as filed:

      ☒ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US19/65857

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 4-99, 115, 121, 146, 147, 153, 154, 160, 161, 167-169, 175-177, 201-203, 209-211
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

-***-Please see supplemental page-***-

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-3, 100-114, 116-120, 122-133, 139-145, 148-152, 155-159, 162-166, 170-174, 178-183

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US19/65857

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C12Q1/6813, 1/682, 1/6837, 1/6841; G01N 21/64, 33/533, 33/58 (2020.01)

CPC - C12Q 1/6813, 1/682, 1/6837, 1/6841; G01N 21/6486, 33/533, 33/582

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X – Y	WO 2018/026873 A1 (CALIFORNIA INSTITUTE OF TECHNOLOGY) 8 February 2018; paragraphs [0008], [0015], [0055], [0056], [0057], [00115], [00141], [00155], [00227], [00289], [00294], [00316], [00413]; Fig. 4a-c	1-2, 3/1-2, 100-103, 105-108, 122-133, 141-144, 148-151, 155-158, 178-182 104/102-103, 109/107-108, 110-113, 114/112-113, 116-119, 120/118-119, 139-140, 145/143-144, 152/150-151, 159/157-158, 162-165, 166/164-165, 170-173, 174/172-173, 183/181-182
Y	(ZHANG, J et al.) Tandem Spinach Array for mRNA Imaging in Living Bacterial Cells. Scientific Reports. 27 November 2015, Vol. 5, No. 17295; pages 1-9; abstract; page 2, 4th paragraph; Fig. 1C; page DOI: 10.1038/srep17295	104/102-103, 109/107-108, 114/112-113, 120/118-119, 145/143-144, 152/150-151, 159/157-158, 166/164-165, 174/172-17, 183/181-182

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
25 March 2020 (25.03.2020)	01 MAY 2020
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Shane Thomas Telephone No. PCT Helpdesk: 571-272-4300

Form PCT/ISA/210 (second sheet) (July 2019)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US19/65857

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2018/089438 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 17 May 2018; page 20, lines 1-5	110-113, 114/112-113, 116-119, 120/118-119, 139-140, 162-165, 166/164-165, 170-173, 174/172-173
A	(CHEN, KH et al.) RNA Imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science. 24 April 2015, Epub 9 April 2015, Vol. 348, No. 6233; pages 1-14; DOI: 10.1126/science.aaa6090	1-2, 3/1-2, 100-103, 104/102-103, 105-108, 109/107-108, 110-113, 114/112-113, 116-119, 120/118-119, 122-133, 139-144, 145/143-144, 148-151, 152/150-151, 155-158, 159/157-158, 162-165, 166/164-165, 170-173, 174/172-173, 178-182, 183/181-182
A	US 2017/0220733 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 3 August 2017; entire document	1-2, 3/1-2, 100-103, 104/102-103, 105-108, 109/107-108, 110-113, 114/112-113, 116-119, 120/118-119, 122-133, 139-144, 145/143-144, 148-151, 152/150-151, 155-158, 159/157-158, 162-165, 166/164-165, 170-173, 174/172-173, 178-182, 183/181-182

Form PCT/ISA/210 (continuation of second sheet) (July 2019)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US19/65857

-***-Continued from Box No. III: Observations where unity of invention is lacking-***-

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, Claims 1-3, 106-114, 116-120, 122-133, 139-145, 148-152, 155-159, 162-166, 170-174, and 178-183 are directed toward methods comprising exposing a sample to nucleic acid probes, amplifying the probes, determining a distribution of the probes within the sample, creating codewords based on said distribution, and error correcting invalid codewords.

Group II, Claims 134-138 are directed toward a method comprising exposing a sample to nucleic acid probes, binding amplifying nucleic acids within a fixed distance to the probes, determining a distribution of the probes within the sample, and creating codewords based on said distribution.

Group III, Claims 164-200 and 204-208 are directed toward methods comprising exposing a sample to a binding or targeting entity conjugated to nucleic acid probes, binding nucleic acids to the probes, determining a distribution of the probes within the sample, and creating codewords based on said distribution.

The inventions listed as Groups I- and III do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical features of Group I include error correction, not present in either of Groups II or III; the special technical features of Group II include primary amplifier nucleic acids which bind to probes within a fixed distance, not present in either of Groups I or III; the special technical features of Group III include a targeting or binding entity, not present in either of Groups I or II.

Groups I-III share the technical features including: a method, comprising: exposing a sample to nucleic acid probes; exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes; determining a distribution of the nucleic acid probes within the sample using fluorescence; and creating codewords based on the fluorescence distribution within the sample. Groups I and III further share the technical features including: wherein the binding of primary amplifier nucleic acids and secondary amplifier nucleic acids to a target is saturatable; exposing primary amplifier nucleic acids to secondary amplifier nucleic acids able to bind to the primary amplifier nucleic acids; wherein the primary amplifier nucleic acids are formed from only 3 of the 4 naturally-occurring nucleotides; and wherein the secondary amplifier nucleic acids are formed from only 3 of the 4 naturally-occurring nucleotides.

However, these shared technical features are previously disclosed by WO 2016/026873 A1 to California Institute of Technology (hereinafter 'California') in view of US 6,150,516 A to Brenner et al. (hereinafter 'Brenner').

California discloses a method, comprising: exposing a sample to nucleic acid probes (a method, comprising: exposing a sample to nucleic acid probes; paragraph [0008]); exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes (exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes; paragraphs [00115], [00125], [00227]); determining a distribution of the nucleic acid probes within the sample using fluorescence (visually identifying probes bound to targets within a sample (determining a distribution of the nucleic acid probes within the sample using fluorescence); paragraph [0008]); and creating codewords based on the fluorescence distribution within the sample (and barcoding the sample based on serial probe hybridization and visualization rounds (creating codewords based on the fluorescence distribution within the sample); paragraph [0006]); wherein the binding of primary amplifier nucleic acids and secondary amplifier nucleic acids to a target is saturatable (wherein there are N molecular targets in the sample (wherein the binding of primary amplifier nucleic acids and secondary amplifier nucleic acids to a target is saturatable); paragraphs [0008], [00115], [00227]; wherein the limited number of targets naturally creates a saturatable level of target binding sites for probes and amplifier nucleic acids); and exposing primary amplifier nucleic acids to secondary amplifier nucleic acids able to bind to the primary amplifier nucleic acids (exposing secondary probes (primary amplifier nucleic acids) to tertiary probes (secondary amplifier nucleic acids) able to bind to the secondary probes (primary amplifier nucleic acids); paragraphs [00115], [00125], [00227]).

California does not disclose: wherein the primary amplifier nucleic acids are formed from only 3 of the 4 naturally-occurring nucleotides; and wherein the secondary amplifier nucleic acids are formed from only 3 of the 4 naturally-occurring nucleotides.

Brenner discloses wherein minimally cross-hybridizing sets of oligonucleotide tags and complements thereof may be synthesized (wherein minimally cross-hybridizing sets of oligonucleotide tags and complements thereof may be synthesized; column 3, lines 36-40), depending on the size of the set desired and the degree to which cross-hybridization is sought to be minimized (depending on the size of the set desired and the degree to which cross-hybridization is sought to be minimized; column, 3, lines 36-40), using 3 of the 4 different types of nucleotides (using 3 of the 4 different types of nucleotides; column 3, lines 47-47, lines 56-59).

It would have been obvious to a person of ordinary skill in the art at the time the invention was made to have modified the disclosure of California to have used only 3 of the 4 different types of naturally occurring nucleotides for the formation of minimally cross-hybridizing sets of probes, based on the disclosure of Brenner, in order to better enable multiplexing of the detection of targets in a sample using a plurality of non-cross-hybridizing amplification probes, based on the disclosure of California, to more efficiently and effectively detect and barcode target sequences within the sample.

Since none of the special technical features of the Groups I-III inventions is found in more than one of the inventions, and since all of the shared technical features are previously disclosed by a combination of the California and Brenner references, unity of invention is lacking.

RECEIVED
Wolf, Greenfield & Sacks, P.C.
MAY 04 2020

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

Docketed ____ Already Docketed ____
Not Required ____
Initials 1st ____ 2nd ____

To: Tani Chen
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
United States of America

PCT

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing *(day/month/year)*	01 MAY 2020

Applicant's or agent's file reference	FOR FURTHER ACTION
H0498.70680WO00	See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US19/65857	12 December 2019 (12.12.2019)	13 December 2018 (13.12.2018)

International Patent Classification (IPC) or both national classification and IPC
IPC - C12Q1/6813, 1/682, 1/6837, 1/6841; G01N 21/64, 33/533, 33/58 (2020.01)
CPC - C12Q 1/6813, 1/682, 1/6837, 1/6841; G01N 21/6486, 33/533, 33/582

Applicant
PRESIDENT AND FELLOWS OF HARVARD COLLEGE

1. This opinion contains indications relating to the following items:

- [X] Box No. I Basis of the opinion
- [ ] Box No. II Priority
- [X] Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- [X] Box No. IV Lack of unity of invention
- [X] Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
- [ ] Box No. VI Certain documents cited
- [ ] Box No. VII Certain defects in the international application
- [ ] Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA/US	Date of completion of this opinion	Authorized officer
Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	25 March 2020 (25.03.2020)	Shane Thomas PCT Help Desk Telephone No. 571-272-4300

Form PCT/ISA/237 (cover sheet) (revised January 2019)

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US19/65857

Box No. I Basis of this opinion

1. With regard to the **language**, this opinion has been established on the basis of:

   ☒ the international application in the language in which it was filed.

   ☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(b)).

3. ☒ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:

   a. ☒ forming part of the international application as filed:

      ☒ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

Form PCT/ISA/237 (Box No. I) (revised January 2019)

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US19/65857

Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☒ claims Nos. 4-99, 115, 121, 146, 147, 153, 154, 160, 161, 167-169, 175-177, 201-203, 209-211

because:

☐ the said international application, or the said claims Nos. _____ relate to the following subject matter which does not require an international search *(specify)*:

☒ the description, claims or drawings *(indicate particular elements below)* or said claims Nos. -***-Please see below-***- are so unclear that no meaningful opinion could be formed *(specify)*:

Claims Nos. 4-99, 115, 121, 146, 147, 153, 154, 160, 161, 167-169, 175-177, 201-203, and 209-211 are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

☐ the claims, or said claims Nos. _____ are so inadequately supported by the description that no meaningful opinion could be formed *(specify)*:

☒ no international search report has been established for said claims Nos. -***-Please see above-***-

☐ a meaningful opinion could not be formed without the sequence listing; the applicant did not, within the prescribed time limit:

- ☐ furnish a sequence listing in the form of an Annex C/ST.25 text file, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.
- ☐ furnish a sequence listing on paper or in the form of an image file complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.
- ☐ pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rule 13*ter*.1(a) or (b).

☐ See Supplemental Box for further details.

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US19/65857

Box No. IV Lack of unity of invention

1. ☒ In response to the invitation (Form PCT/ISA/206) to pay additional fees the applicant has, within the applicable time limit:

☐ paid additional fees.

☐ paid additional fees under protest and, where applicable, the protest fee.

☐ paid additional fees under protest but the applicable protest fee was not paid.

☒ not paid additional fees.

2. ☐ This Authority found that the requirement of unity of invention is not complied with and chose not to invite the applicant to pay additional fees.

3. This Authority considers that the requirement of unity of invention in accordance with Rule 13.1, 13.2 and 13.3 is

☐ complied with.

☒ not complied with for the following reasons:

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, Claims 1-3, 100-114, 116-120, 122-133, 139-145, 148-152, 155-159, 162-166, 170-174, and 178-183 are directed toward methods comprising exposing a sample to nucleic acid probes, amplifying the probes, determining a distribution of the probes within the sample, creating codewords based on said distribution, and error correcting invalid codewords.

Group II, Claims 134-138 are directed toward a method comprising exposing a sample to nucleic acid probes, binding amplifying nucleic acids within a fixed distance to the probes, determining a distribution of the probes within the sample, and creating codewords based on said distribution.

Group III, Claims 184-200 and 204-208 are directed toward methods comprising exposing a sample to a binding or targeting entity conjugated to nucleic acid probes, binding nucleic acids to the probes, determining a distribution of the probes within the sample, and creating codewords based on said distribution.

The inventions listed as Groups I- and III do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical features of Group I include error correction, not present in either of Groups II or III; the special technical features of Group II include primary amplifier nucleic acids which bind to probes within a fixed distance, not present in either of Groups I or III; the special technical features of Group III include a targeting or binding entity, not present in either of Groups I or II.

Groups I-III share the technical features including: a method, comprising: exposing a sample to nucleic acid probes; exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes; determining a distribution of the nucleic acid probes within the sample using fluorescence; and creating codewords based on the fluorescence distribution within the sample. Groups I and III further share the technical features including: wherein the binding of primary amplifier nucleic acids and secondary amplifier nucleic acids to a target is saturatable; exposing primary amplifier nucleic acids to secondary amplifier nucleic acids able to bind to the primary amplifier nucleic acids; wherein the primary amplifier nucleic acids are formed from only 3 of the 4 naturally-occuring nucleotides; and wherein the secondary amplifier nucleic acids are formed from only 3 of the 4 naturally-occurring nucleotides.

However, these shared technical features are previously disclosed by WO 2018/026873 A1 to California Institute of Technology (hereinafter 'California') in view of US 6,150,516 A to Brenner et al. (hereinafter 'Brenner').

-***-Continued on supplemental page-***-

4. Consequently, this opinion has been established in respect of the following parts of the international application:

☐ all parts.

☒ the parts relating to claims Nos. 1-3, 100-114, 116-120, 122-133, 139-145, 148-152, 155-159, 162-166, 170-174,178-183

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US19/65857

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	-***-Please see below-***-	YES
	Claims	-***-Please see below-***-	NO
Inventive step (IS)	Claims	NONE	YES
	Claims	-***-Please see below-***-	NO
Industrial applicability (IA)	Claims	-***-Please see below-***-	YES
	Claims	NONE	NO

2. Citations and explanations:

-***-Continued from Box No. V: 1. Statement-***-

Novelty (N): YES: Claims 104/102-103, 109/107-108, 110-113, 114/112-113, 116-119, 120/116-119, 139-140, 145/143-144, 152/150-151, 159/157-158, 162-165, 166/164-165, 170-173, 174/172-173, 183/181-182

Novelty (N): NO: Claims 1-2, 3/1-2, 100-103, 105-108, 122-133, 141-144, 148-151, 155-158, 178-182

Inventive step (IS): NO: Claims 1-2, 3/1-2, 100-103, 104/102-103, 105-108, 109/107-108, 110-113, 114/112-113, 116-119, 120/116-119, 122-133, 139-144, 145/143-144, 148-151, 152/150-151, 155-158, 159/157-158, 162-165, 166/164-165, 170-173, 174/172-173, 178-182, 183/181-182

Industrial Applicability (IA): YES: Claims 1-2, 3/1-2, 100-103, 104/102-103, 105-108, 109/107-108, 110-113, 114/112-113, 116-119, 120/116-119, 122-133, 139-144, 145/143-144, 148-151, 152/150-151, 155-158, 159/157-158, 162-165, 166/164-165, 170-173, 174/172-173, 178-182, 183/181-182

-***-Box No. V: 2. Citations and Explanations-***-

Claims 1-2, 3/1-2, 100-103, 105-108, 122-133, 141-144, 148-151, 155-158, and 178-182 lack novelty under PCT Article 33(2) as being anticipated by WO 2018/026873 A1 to California Institute of Technology (hereinafter 'California').

As per claim 1, California discloses a method, comprising: exposing a sample to nucleic acid probes (a method of barcoding molecular targets comprising a plurality of primary probes, wherein each probe comprises a primary binding sequence that binds to a [illegible] complementary target sequence (exposing) in a target nucleic acid molecule (sample); paragraphs [0008], [0055]); exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes (a plurality of bridge probes, wherein each bridge probe (primary amplifier nucleic acid) comprises a binding sequence that specifically binds to the overhang sequence of a primary probe, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0056], [00115], [00227], [00289], [00294]), wherein a maximum number of primary amplifier nucleic acids is able to bind to a nucleic acid probe (two or more barcoding rounds, wherein each barcoding round comprises 2 or more serial probe hybridizations; in which secondary probes (primary amplifier nucleic acids) bind to the primary probes, collectively bound to N molecular targets, thereby indicating that a maximum number of primary amplifier probes bind the nucleic acid probe; paragraphs [0008], [0056], [00115], [00227], [00289], [00294]); exposing the primary amplifier nucleic acids to secondary amplifier nucleic acids able to bind to the primary amplifier nucleic acids (tertiary probes (secondary nucleic acids) that bind to secondary probes (primary amplifier nucleic acids) that bind to primary probes that directly bind to a target nucleic acid molecule, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0055], [0056], [0057], [00115], [00227], [00289], [00294]), wherein a maximum number of secondary amplifier nucleic acids is able to bind to the primary amplifier nucleic acids (two or more barcoding rounds, wherein each barcoding round comprises 2 or more serial probe hybridizations, in which tertiary probes (secondary amplifier nucleic acids) bind to secondary probes that bind to primary probes, collectively bound to N molecular targets, thereby indicating that a maximum number of secondary amplifier probes bind the primary amplifier probe; paragraphs [0008], [0056], [00115], [00227], [00289], [00294]); determining a distribution of the nucleic acid probes within the sample using fluorescence (probes generate one or more different detectable visual fluorescent signals corresponding to the number of molecular targets (determining a distribution of probes); paragraphs [0008], [00141]); creating codewords based on the fluorescence distribution within the sample (barcoding the sample based on serial probe hybridization and visualization rounds, wherein the probes generate a detectable fluorescent signal (creating codewords based on the fluorescence distribution within the sample); paragraphs [0008], [00141]); and for at least some of the codewords, matching the codeword to a valid codeword (data from repeat rounds can validate each other, for example, a cyan data point in an image corresponding to hybridization round 2 can be matched to the image corresponding to hybridization round 3 at the same location, revealing a yellow data point, and identifying the data point as part of mRNA1 (valid codeword); paragraph [00316]) optionally wherein, if no match is found, applying error correction to the codewords to form a valid codeword (the barcoding rounds include one or more rounds of error correction, which would intrinsically be applied to the barcoding if a valid match is not found, in order to form a valid code; paragraph [0015]).

-***-Continued Within the Next Supplemental Box-***-

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

-***-Continued from Box V: Citations and Explanations-***-

As per claim 2, California discloses the method of claim 1, and California further discloses comprising exposing the sample to at least 5 different nucleic acid probes (RNA molecules are captured on LNA poly-dT functionalized coverslips followed by hybridization of 28-32 gene specific primary probes; paragraph [00155]; Fig. 4a-c).

As per claims 3/1-2, California discloses the method of claims 1-2, and California further discloses comprising exposing the sample to at least 10 different nucleic acid probes (RNA molecules are captured on LNA poly-dT functionalized coverslips followed by hybridization of 28-32 gene specific primary probes; paragraph [00155]; Fig. 4a-c).

As per claim 100, California discloses a method, comprising: exposing a sample to nucleic acid probes (a method of barcoding molecular targets comprising a plurality of primary probes, wherein each probe comprises a primary binding sequence that binds to a complementary target sequence (exposing) in a target nucleic acid molecule (sample); paragraphs [0008], [0055]); exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes (a plurality of bridge probes, wherein each bridge probe (primary amplifier nucleic acid) comprises a binding sequence that specifically binds to the overhang sequence of a primary probe, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0056], [00115], [00227], [00289], [00294]); exposing the primary amplifier nucleic acids to secondary amplifier nucleic acids able to bind to the primary amplifier nucleic acids (tertiary probes (secondary nucleic acids) that bind to secondary probes (primary amplifier nucleic acids) that bind to primary probes that directly bind to a target nucleic acid molecule, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0055], [0056], [0057], [00115], [00227], [00289], [00294]), wherein the binding of primary amplifier nucleic acids and secondary amplifier nucleic acids to a target is saturatable (two or more barcoding rounds, wherein each barcoding round comprises 2 or more serial probe hybridizations, in which tertiary probes (secondary amplifier nucleic acids) bind to secondary probes (primary amplifier nucleic acids), which in turn bind to the primary probes, collectively bound to N molecular targets, thereby indicating that the binding of primary amplifier nucleic acids and secondary amplifier nucleic acids to a target is saturatable; paragraphs [0008], [0056], [00115], [00227], [00289], [00294]); determining a distribution of the nucleic acid probes within the sample using fluorescence (probes generate one or more different detectable visual fluorescent signals corresponding to the number of molecular targets (determining a distribution of probes); paragraphs [0008], [00141]); creating codewords based on the fluorescence distribution within the sample (barcoding the sample based on serial probe hybridization and visualization rounds, wherein the probes generate a detectable fluorescent signal (creating codewords based on the fluorescence distribution within the sample); paragraphs [0008], [00141]); and for at least some of the codewords, matching the codeword to a valid codeword (data from repeat rounds can validate each other, for example, a cyan data point in an image corresponding to hybridization round 2 can be matched to the image corresponding to hybridization round 3 at the same location, revealing a yellow data point, and identifying the data point as part of mRNA1 (valid codeword); paragraph [00316]) optionally wherein, if no match is found, applying error correction to the codeword to form a valid codeword (the barcoding rounds include one or more rounds of error correction, which would intrinsically be applied to the barcoding if a valid match is not found, in order to form a valid code; paragraph [0015]).

As per claim 101, California discloses the method of claim 100, and California further discloses comprising exposing the secondary amplifier nucleic acids to one or more additional rounds of amplifier nucleic acids, including an ending round of amplifier nucleic acids, each able to bind the amplifier nucleic acids of the previous round (five populations of extender probes, each binding to a region of the probe population of the previous round, e.g. the second population binds a region in the first population, the third population binds a region in the second population, and so on, can be used to amplify the signal up to 5 times, with the fifth population of probes being the final population (ending round), thereby indicating exposing the tertiary probe (secondary amplifier nucleic acid) to one or more additional rounds of amplifier nucleic acids; paragraphs [0056], [00115], [00227], [00289], [00294]).

As per claim 102, California discloses the method of claim 101, and California further discloses wherein the ending amplifier nucleic acids comprise a fluorescent signaling entity (probes from all rounds, including the ending amplifier nucleic acid, are connected with a signal moiety, such as a fluorophore, that can produce a detectable signal; paragraphs [0056], [00115], [00141], [00227], [00289], [00294]).

As per claim 103, California discloses the method of claim 101, and California further discloses comprising exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity (inserting a cleavable linker between the binding sequence and fluorescent signal moiety of a probe, wherein for the next round of hybridization (ending round of amplifier nucleic acids), the visible signals are switched, thereby indicating exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity; paragraphs [0056], [00115], [00141], [00227], [00227], [00289], [00294]).

-***-Continued Within the Next Supplemental Box-***-

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

-***-Continued from Previous Supplemental Box-***-

As per claim 105, California discloses a method, comprising: exposing a sample to nucleic acid probes (a method of barcoding molecular targets comprising a plurality of primary probes, wherein each probe comprises a primary binding sequence that binds to a complementary target sequence (exposing) in a target nucleic acid molecule (sample); paragraphs [0008], [0055]); exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes (a plurality of bridge probes, wherein each bridge probe (primary amplifier nucleic acid) comprises a binding sequence that specifically binds to the overhang sequence of a primary probe, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0056], [00115], [00227], [00289], [00294]); exposing the primary amplifier nucleic acids to secondary amplifier nucleic acids able to bind to the primary amplifier nucleic acids (tertiary probes (secondary nucleic acids) that bind to secondary probes (primary amplifier nucleic acids) that bind to primary probes that directly bind to a target nucleic acid molecule, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0055], [0056], [0057], [00115], [00227], [00289], [00294]), wherein the secondary amplifier nucleic acids bind to the primary amplifier nucleic acids within a fixed distance (false positive signals were detected on barcodes hamming distance one away from on-target barcodes, thereby indicating that the tertiary probes (secondary amplifier nucleic acids) bind to the secondary probes (primary amplifier nucleic acids) within a fixed distance; paragraphs [0055], [0056], [0057], [00115], [00227], [00289], [00294], [00413]); determining a distribution of the nucleic acid probes within the sample using fluorescence (probes generate one or more different detectable visual fluorescent signals corresponding to the number of molecular targets (determining a distribution of probes); paragraphs [0008], [00141]); creating codewords based on the fluorescence distribution within the sample (barcoding the sample based on serial probe hybridization and visualization rounds, wherein the probes generate a detectable fluorescent signal (creating codewords based on the fluorescence distribution within the sample); paragraphs [0008], [00141]); and for at least some of the codewords, matching the codeword to a valid codeword (data from repeat rounds can validate each other, for example, a cyan data point in an image corresponding to hybridization round 2 can be matched to the image corresponding to hybridization round 3 at the same location, revealing a yellow data point, and identifying the data point as part of mRNA1 (valid codeword); paragraph [00316]) optionally wherein, if no match is found, applying error correction to the codeword to form a valid codeword (the barcoding rounds include one or more rounds of error correction, which would intrinsically be applied to the barcoding if a valid match is not found, in order to form a valid code; paragraph [0015]).

As per claim 106, California discloses the method of claim 105, and California further discloses comprising exposing the secondary amplifier nucleic acids to one or more additional rounds of amplifier nucleic acids, including an ending round of amplifier nucleic acids, each able to bind the amplifier nucleic acids of the previous round (five populations of extender probes, each binding to a region of the probe population of the previous round, e.g. the second population binds a region in the first population, the third population binds a region in the second population, and so on, can be used to amplify the signal up to 5 times, with the fifth population of probes being the final population (ending round), thereby indicating exposing the tertiary probe (secondary amplifier nucleic acid) to one or more additional rounds of amplifier nucleic acids; paragraphs [0056], [00115], [00227], [00289], [00294]).

As per claim 107, California discloses the method of claim 106, and California further discloses wherein the ending amplifier nucleic acids comprise a fluorescent signaling entity (probes from all rounds, including the ending amplifier nucleic acid, are connected with a signal moiety, such as a fluorophore, that can produce a detectable signal; paragraphs [0056], [00115], [00141], [00227], [00289], [00294]).

As per claim 108, California discloses the method of claim 105, and California further discloses comprising exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity (inserting a cleavable linker between the binding sequence and fluorescent signal moiety of a probe, wherein for the next round of hybridization (ending round of amplifier nucleic acids), the visible signals are switched, thereby indicating exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity; paragraphs [0056], [00115], [00141], [00227], [00227], [00289], [00294]).

As per claim 122, California discloses a method, comprising: exposing a sample to nucleic acid probes (a method of barcoding molecular targets comprising a plurality of primary probes, wherein each probe comprises a primary binding sequence that binds to a complementary target sequence (exposing) in a target nucleic acid molecule (sample); paragraphs [0008], [0055]); exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes (a plurality of bridge probes, wherein each bridge probe (primary amplifier nucleic acid) comprises a binding sequence that specifically binds to the overhang sequence of a primary probe, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0056], [00115], [00227], [00289], [00294]), wherein a maximum number of primary amplifier nucleic acids is able to bind to a nucleic acid probe (two or more barcoding rounds, wherein each barcoding round comprises 2 or more serial probe hybridizations, in which secondary probes (primary amplifier probes) bind to the primary probes, collectively bound to N molecular targets, thereby indicating that a maximum number of primary amplifier probes bind the nucleic acid probe; paragraphs [0008], [0056], [00115], [00227], [00289], [00294]); determining a distribution of the nucleic acid probes within the sample using fluorescence (probes generate one or more different detectable visual fluorescent signals corresponding to the number of molecular targets (determining a distribution of probes); paragraphs [0008], [00141]); and creating codewords based on the fluorescence distribution within the sample (barcoding the sample based on serial probe hybridization and visualization rounds, wherein the probes generate a detectable fluorescent signal (creating codewords based on the fluorescence distribution within the sample); paragraphs [0008], [00141]).

As per claim 123, California discloses the method of claim 122, and California further discloses comprising exposing the primary amplifier nucleic acids to a fluorescent signaling entity (inserting a cleavable linker between the binding sequence and fluorescent signal moiety of a probe, wherein for the next round of hybridization (primary amplifier nucleic acids), the visible signals are switched, thereby indicating exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity; paragraphs [0056], [00115], [00141], [00227], [00227], [00289], [00294]).

-***-Continued Within the Next Supplemental Box-***-

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

-***-Continued from Previous Supplemental Box-***-

As per claim 124, California discloses the method of claim 122, and California further discloses wherein the primary amplifier nucleic acids comprise a fluorescent signaling entity (probes including the secondary probes (primary amplifier nucleic acids), are connected with a signal moiety, such as a fluorophore, that can produce a detectable signal; paragraphs [0056], [00115], [00141], [00227], [00289], [00294]).

As per claim 125, California discloses a method, comprising: exposing a sample to nucleic acid probes (a method of barcoding molecular targets comprising a plurality of primary probes, wherein each probe comprises a primary binding sequence that binds to a complementary target sequence (exposing) in a target nucleic acid molecule (sample); paragraphs [0008], [0055]); exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes (a plurality of bridge probes, wherein each bridge probe (primary amplifier nucleic acid) comprises a binding sequence that specifically binds to the overhang sequence of a primary probe, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0056], [00115], [00227], [00289], [00294]), wherein a maximum number of primary amplifier nucleic acids is able to bind to a nucleic acid probe (two or more barcoding rounds, wherein each barcoding round comprises 2 or more serial probe hybridizations, in which secondary probes (primary amplifier probes) bind to the primary probes, collectively bound to N molecular targets, thereby indicating that a maximum number of primary amplifier probes bind the nucleic acid probe; paragraphs [0008], [0056], [00115], [00227], [00289], [00294]); determining a distribution of the nucleic acid probes within the sample using fluorescence (probes generate one or more different detectable visual fluorescent signals corresponding to the number of molecular targets (determining a distribution of probes); paragraphs [0008], [00141]); and creating codewords based on the fluorescence distribution within the sample (barcoding the sample based on serial probe hybridization and visualization rounds, wherein the probes generate a detectable fluorescent signal (creating codewords based on the fluorescence distribution within the sample); paragraphs [0008], [00141]), and for at least some of the codewords, matching the codeword to a valid codeword (data from repeat rounds can validate each other, for example, a cyan data point in an image corresponding to hybridization round 2 can be matched to the image corresponding to hybridization round 3 at the same location, revealing a yellow data point, and identifying the data point as part of mRNA1 (valid codeword); paragraph [00316]) wherein, if no match is found, potentially applying error correction to the codeword to form a valid codeword (the barcoding rounds include one or more rounds of error correction, which would intrinsically be applied to the barcoding if a valid match is not found, in order to form a valid code; paragraph [0015]).

As per claim 126, California discloses the method of claim 125, and California further discloses comprising exposing the primary amplifier nucleic acids to a fluorescent signaling entity (inserting a cleavable linker between the binding sequence and fluorescent signal moiety of a probe, wherein for the next round of hybridization (primary amplifier nucleic acids), the visible signals are switched, thereby indicating exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity; paragraphs [0056], [00115], [00141], [00227], [00227], [00289], [00294]).

As per claim 127, California discloses the method of claim 125, and California further discloses wherein the primary amplifier nucleic acids comprise a fluorescent signaling entity (probes including the secondary probes (primary amplifier nucleic acids), are connected with a signal moiety, such as a fluorophore, that can produce a detectable signal; paragraphs [0056], [00115], [00141], [00227], [00289], [00294]).

As per claim 128, California discloses a method, comprising: exposing a sample to nucleic acid probes (a method of barcoding molecular targets comprising a plurality of primary probes, wherein each probe comprises a primary binding sequence that binds to a complementary target sequence (exposing) in a target nucleic acid molecule (sample); paragraphs [0008], [0055]); exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes (a plurality of bridge probes, wherein each bridge probe (primary amplifier nucleic acid) comprises a binding sequence that specifically binds to the overhang sequence of a primary probe, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0056], [00115], [00227], [00289], [00294]), wherein the binding of primary amplifier nucleic acids to a target is saturatable (two or more barcoding rounds, wherein each barcoding round comprises 2 or more serial probe hybridizations, in which tertiary probes bind to secondary probes (primary amplifier nucleic acids), which in turn bind to the primary probes, collectively bound to N molecular targets, or thereby indicating that the binding of primary amplifier nucleic acids to a target is saturatable; paragraphs [0008], [0056], [00115], [00227], [00289], [00294]); determining a distribution of the nucleic acid probes within the sample using fluorescence (probes generate one or more different detectable visual fluorescent signals corresponding to the number of molecular targets (determining a distribution of probes); paragraphs [0008], [00141]); and creating codewords based on the fluorescence distribution within the sample (barcoding the sample based on serial probe hybridization and visualization rounds, wherein the probes generate a detectable fluorescent signal (creating codewords based on the fluorescence distribution within the sample); paragraphs [0008], [00141]).

As per claim 129, California discloses the method of claim 128, and California further discloses comprising exposing the primary amplifier nucleic acids to a fluorescent signaling entity (inserting a cleavable linker between the binding sequence and fluorescent signal moiety of a probe, wherein for the next round of hybridization (primary amplifier nucleic acids), the visible signals are switched, thereby indicating exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity; paragraphs [0056], [00115], [00141], [00227], [00227], [00289], [00294]).

As per claim 130, California discloses the method of claim 128, and California further discloses wherein the primary amplifier nucleic acids comprise a fluorescent signaling entity (probes including the secondary probes (primary amplifier nucleic acids), are connected with a signal moiety, such as a fluorophore, that can produce a detectable signal; paragraphs [0056], [00115], [00141], [00227], [00289], [00294]).

-***-Continued Within the Next Supplemental Box-***-

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

-***-Continued from Previous Supplemental Box-***-

As per claim 131, California discloses a method, comprising: exposing a sample to nucleic acid probes (a method of barcoding molecular targets comprising a plurality of primary probes, wherein each probe comprises a primary binding sequence that binds to a complementary target sequence (exposing) in a target nucleic acid molecule (sample); paragraphs [0008], [0055]); exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes (a plurality of bridge probes, wherein each bridge probe (primary amplifier nucleic acid) comprises a binding sequence that specifically binds to the overhang sequence of a primary probe, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0056], [00115], [00227], [00289], [00294]), wherein the binding of primary amplifier nucleic acids to a target is saturatable (two or more barcoding rounds, wherein each barcoding round comprises 2 or more serial probe hybridizations, in which tertiary probes bind to secondary probes (primary amplifier nucleic acids), which in turn bind to the primary probes, collectively bound to N molecular targets, thereby indicating that the binding of primary amplifier nucleic acids to a target is saturatable; paragraphs [0008], [0056], [00115], [00227], [00289], [00294]); determining a distribution of the nucleic acid probes within the sample using fluorescence (probes generate one or more different detectable visual fluorescent signals corresponding to the number of molecular targets (determining a distribution of probes); paragraphs [0008], [00141]); and creating codewords based on the fluorescence distribution within the sample (barcoding the sample based on serial probe hybridization and visualization rounds, wherein the probes generate a detectable fluorescent signal (creating codewords based on the fluorescence distribution within the sample); paragraphs [0008], [00141]), and for at least some of the codewords, matching the codeword to a valid codeword (data from repeat rounds can validate each other, for example, a cyan data point in an image corresponding to hybridization round 2 can be matched to the image corresponding to hybridization round 3 at the same location, revealing a yellow data point, and identifying the data point as part of mRNA1 (valid codeword); paragraph [00316]) wherein, if no match is found, potentially applying error correction to the codeword to form a valid codeword (the barcoding rounds include one or more rounds of error correction, which would intrinsically be applied to the barcoding if a valid match is not found in order to form a valid code; paragraph [00315]).

As per claim 132, California discloses the method of claim 131, and California further discloses comprising exposing the primary amplifier nucleic acids to a fluorescent signaling entity (inserting a cleavable linker between the binding sequence and fluorescent signal moiety of a probe, wherein for the next round of hybridization (primary amplifier nucleic acids), the visible signals are switched, thereby indicating exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity; paragraphs [0056], [00115], [00141], [00227], [00227], [00289], [00294]).

As per claim 133, California discloses the method of claim 131, and California further discloses wherein the primary amplifier nucleic acids comprise a fluorescent signaling entity (probes including the secondary probes (primary amplifier nucleic acids), are connected with a signal moiety, such as a fluorophore, that can produce a detectable signal; paragraphs [0056], [00115], [00141], [00227], [00289], [00294]).

As per claim 141, California discloses a method, comprising: exposing a sample to nucleic acid probes (a method of barcoding molecular targets comprising a plurality of primary probes, wherein each probe comprises a primary binding sequence that binds to a complementary target sequence (exposing) in a target nucleic acid molecule (sample); paragraphs [0008], [0055]); exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes (a plurality of bridge probes, wherein each bridge probe (primary amplifier nucleic acid) comprises a binding sequence that specifically binds to the overhang sequence of a primary probe, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0056], [00115], [00227], [00289], [00294]), wherein a maximum number of primary amplifier nucleic acids is able to bind to a nucleic acid probe (two or more barcoding rounds, wherein each barcoding round comprises 2 or more serial probe hybridizations, in which secondary probes (primary amplifier probes) bind to the primary probes, collectively bound to N molecular targets, thereby indicating that a maximum number of primary amplifier probes bind the nucleic acid probe; paragraphs [0008], [0056], [00115], [00227], [00289], [00294]); exposing the primary amplifier nucleic acids to secondary amplifier nucleic acids able to bind to the primary amplifier nucleic acids (tertiary probes (secondary nucleic acids) that bind to secondary probes (primary amplifier nucleic acids) that bind to primary probes that directly bind to a target nucleic acid molecule, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0055], [0056], [0057], [00115], [00227], [00289], [00294]), wherein a maximum number of secondary amplifier nucleic acids is able to bind to the primary amplifier nucleic acids (two or more barcoding rounds, wherein each barcoding round comprises 2 or more serial probe hybridizations, in which tertiary probes (secondary amplifier nucleic acids) bind to secondary probes that bind to primary probes, collectively bound to N molecular targets, thereby indicating that a maximum number of secondary amplifier probes bind the primary amplifier probe; paragraphs [0008], [0056], [00115], [00227], [00289], [00294]); determining a distribution of the nucleic acid probes within the sample using fluorescence (probes generate one or more different detectable visual fluorescent signals corresponding to the number of molecular targets (determining a distribution of probes); paragraphs [0008], [00141]); and creating codewords based on the fluorescence distribution within the sample (barcoding the sample based on serial probe hybridization and visualization rounds, wherein the probes generate a detectable fluorescent signal (creating codewords based on the fluorescence distribution within the sample); paragraphs [0008], [00141]).

As per claim 142, California discloses the method of claim 141, and California further discloses comprising exposing the secondary amplifier nucleic acids to one or more additional rounds of amplifier nucleic acids, including an ending round of amplifier nucleic acids, each able to bind the amplifier nucleic acids of the previous round (five populations of extender probes, each binding to a region of the probe population of the previous round, e.g. the second population binds a region in the first population, the third population binds a region in the second population, and so on, can be used to amplify the signal up to 8 times, with the fifth population of probes being the final population (ending round), thereby indicating exposing the tertiary probe (secondary amplifier nucleic acid) to one or more additional rounds of amplifier nucleic acids; paragraphs [0056], [00115], [00227], [00289], [00294]).

-***-Continued Within the Next Supplemental Box-***-

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

-***-Continued from Previous Supplemental Box-***-

As per claim 143, California discloses the method of claim 142, and California further discloses wherein the ending amplifier nucleic acids comprise a fluorescent signaling entity (probes from all rounds, including the ending amplifier nucleic acid, are connected with a signal moiety, such as a fluorophore, that can produce a detectable signal; paragraphs [0056], [00115], [00141], [00227], [00289], [00294]).

As per claim 144, California discloses the method of claim 142, and California further discloses comprising exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity (inserting a cleavable linker between the binding sequence and fluorescent signal moiety of a probe, wherein for the next round of hybridization (ending round of amplifier nucleic acids), the visible signals are switched, thereby indicating exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity; paragraphs [0056], [00115], [00141], [00227], [00227], [00289], [00294]).

As per claim 148, California discloses a method, comprising: exposing a sample to nucleic acid probes (a method of barcoding molecular targets comprising a plurality of primary probes, wherein each probe comprises a primary binding sequence that binds to a complementary target sequence (exposing) in a target nucleic acid molecule (sample); paragraphs [0008], [0055]); exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes (a plurality of bridge probes, wherein each bridge probe (primary amplifier nucleic acid) comprises a binding sequence that specifically binds to the overhang sequence of a primary probe, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0056], [00115], [00227], [00289], [00294]); exposing the primary amplifier nucleic acids to secondary amplifier nucleic acids able to bind to the primary amplifier nucleic acids (tertiary probes (secondary nucleic acids) that bind to secondary probes (primary amplifier nucleic acids) that bind to primary probes that directly bind to a target nucleic acid molecule, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0055], [0056], [0057], [00115], [00227], [00289], [00294]), wherein the binding of primary amplifier nucleic acids and secondary amplifier nucleic acids to a target is saturatable (two or more barcoding rounds, wherein each barcoding round comprises 2 or more serial probe hybridizations, in which tertiary probes (secondary amplifier nucleic acids) bind to secondary probes (primary amplifier nucleic acids), which in turn bind to the primary probes, collectively bound to N molecular targets, thereby indicating that the binding of primary amplifier nucleic acids and secondary amplifier nucleic acids to a target is saturatable; paragraphs [0008], [0056], [00115], [00227], [00289], [00294]); determining a distribution of the nucleic acid probes within the sample using fluorescence (probes generate one or more different detectable visual fluorescent signals, corresponding to the number of molecular targets (determining a distribution of probes); paragraphs [0008], [0014]]); and creating codewords based on the fluorescence distribution within the sample (barcoding the sample based on serial probe hybridization and visualization rounds, wherein the probes generate a detectable fluorescent signal (creating codewords based on the fluorescence distribution within the sample); paragraphs [0008], [00141]).

As per claim 149, California discloses the method of claim 148, and California further discloses comprising exposing the secondary amplifier nucleic acids to one or more additional rounds of amplifier nucleic acids, including an ending round of amplifier nucleic acids, each able to bind the amplifier nucleic acids of the previous round (five populations of extender probes, each binding to a region of the probe population of the previous round, e.g. the second population binds a region in the first population, the third population binds a region in the second population, and so on, can be used to amplify the signal up to 5 times, with the fifth population of probes being the final population (ending round), thereby indicating exposing the tertiary probe (secondary amplifier nucleic acid) to one or more additional rounds of amplifier nucleic acids; paragraphs [0056], [00115], [00227], [00289], [00294]).

As per claim 150, California discloses the method of claim 149, and California further discloses wherein the ending amplifier nucleic acids comprise a fluorescent signaling entity (probes from all rounds, including the ending amplifier nucleic acid, are connected with a signal moiety, such as a fluorophore, that can produce a detectable signal; paragraphs [0056], [00115], [00141], [00227], [00289], [00294]).

As per claim 151, California discloses the method of claim 149, and California further discloses comprising exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity (inserting a cleavable linker between the binding sequence and fluorescent signal moiety of a probe, wherein for the next round of hybridization (ending round of amplifier nucleic acids), the visible signals are switched, thereby indicating exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity; paragraphs [0056], [00115], [00141], [00227], [00227], [00289], [00294]).

-***-Continued Within the Next Supplemental Box-***-

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US19/65857

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

-***-Continued from Previous Supplemental Box-***-

As per claim 155, California discloses a method, comprising: exposing a sample to nucleic acid probes (a method of barcoding molecular targets comprising a plurality of primary probes, wherein each probe comprises a primary binding sequence that binds to a complementary target sequence (exposing) in a target nucleic acid molecule (sample); paragraphs [0008], [0055]); exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes (a plurality of bridge probes, wherein each bridge probe (primary amplifier nucleic acid) comprises a binding sequence that specifically binds to the overhang sequence of a primary probe, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0056], [00115], [00227], [00289], [00294]); exposing the primary amplifier nucleic acids to secondary amplifier nucleic acids able to bind to the primary amplifier nucleic acids (tertiary probes (secondary nucleic acids) that bind to secondary probes (primary amplifier nucleic acids) that bind to primary probes that directly bind to a target nucleic acid molecule, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0055], [0056], [0057], [00115], [00227], [00289], [00294]), wherein the secondary amplifier nucleic acids bind to the primary amplifier nucleic acids within a fixed distance (false positive signals were detected on barcodes hamming distance one away from on-target barcodes, thereby indicating that the tertiary probes (secondary amplifier nucleic acids) bind to the secondary probes (primary amplifier nucleic acids) within a fixed distance; paragraphs [0055], [0056], [0057], [00115], [00227], [00289], [00294], [00413]); determining a distribution of the nucleic acid probes within the sample using fluorescence (probes generate one or more different detectable visual fluorescent signals corresponding to the number of molecular targets (determining a distribution of probes); paragraphs [0008], [00141]); and creating codewords based on the fluorescence distribution within the sample (barcoding the sample based on serial probe hybridization and visualization rounds, wherein the probes generate a detectable fluorescent signal (creating codewords based on the fluorescence distribution within the sample); paragraphs [0008], [00141]).

As per claim 156, California discloses the method of claim 155, and California further discloses comprising exposing the secondary amplifier nucleic acids to one or more additional rounds of amplifier nucleic acids, including an ending round of amplifier nucleic acids, each able to bind the amplifier nucleic acids of the previous round (five populations of extender probes, each binding to a region of the probe population of the previous round, e.g. the second population binds a region in the first population, the third population binds a region in the second population, and so on, can be used to amplify the signal up to 5 times, with the fifth population of probes being the final population (ending round), thereby indicating exposing the tertiary probe (secondary amplifier nucleic acid) to one or more additional rounds of amplifier nucleic acids; paragraphs [0056], [00115], [00227], [00289], [00294]).

As per claim 157, California discloses the method of claim 156, and California further discloses wherein the ending amplifier nucleic acids comprise a fluorescent signaling entity (probes from all rounds, including the ending amplifier nucleic acid, are connected with a signal moiety, such as a fluorophore, that can produce a detectable signal; paragraphs [0056], [00115], [00141], [00227], [00289], [00294]).

As per claim 158, California discloses the method of claim 156, and California further discloses comprising exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity (inserting a cleavable linker between the binding sequence and fluorescent signal moiety of a probe, wherein for the next round of hybridization (ending round of amplifier nucleic acids), the visible signals are switched, thereby indicating exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity; paragraphs [0056], [00115], [00141], [00227], [00227], [00289], [00294]).

As per claim 176, California discloses a method, comprising: exposing a sample to nucleic acid probes (a method of barcoding molecular targets comprising a plurality of primary probes, wherein each probe comprises a primary binding sequence that binds to a complementary target sequence (exposing) in a target nucleic acid molecule (sample); paragraphs [0008], [0055]); exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes (a plurality of bridge probes, wherein each bridge probe (primary amplifier nucleic acid) comprises a binding sequence that specifically binds to the overhang sequence of a primary probe, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0056], [00115], [00227], [00289], [00294]); determining a distribution of the nucleic acid probes within the sample using fluorescence (probes generate one or more different detectable visual fluorescent signals corresponding to the number of molecular targets (determining a distribution of probes); paragraphs [0008], [00141]); and creating codewords based on the fluorescence distribution within the sample (barcoding the sample based on serial probe hybridization and visualization rounds, wherein the probes generate a detectable fluorescent signal (creating codewords based on the fluorescence distribution within the sample); paragraphs [0008], [00141]).

As per claim 179, California discloses a method, comprising: exposing a sample to nucleic acid probes (a method of barcoding molecular targets comprising a plurality of primary probes, wherein each probe comprises a primary binding sequence that binds to a complementary target sequence (exposing) in a target nucleic acid molecule (sample); paragraphs [0008], [0055]); exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes (a plurality of bridge probes, wherein each bridge probe (primary amplifier nucleic acid) comprises a binding sequence that specifically binds to the overhang sequence of a primary probe, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0056], [00115], [00227], [00289], [00294]); exposing the primary amplifier nucleic acids to secondary amplifier nucleic acids able to bind to the primary amplifier nucleic acids (tertiary probes (secondary nucleic acids) that bind to secondary probes (primary amplifier nucleic acids) that bind to primary probes that directly bind to a target nucleic acid molecule, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0055], [0056], [0057], [00115], [00227], [00289], [00294]); determining a distribution of the nucleic acid probes within the sample using fluorescence (probes generate one or more different detectable visual fluorescent signals corresponding to the number of molecular targets (determining a distribution of probes); paragraphs [0008], [00141]); and creating codewords based on the fluorescence distribution within the sample (barcoding the sample based on serial probe hybridization and visualization rounds, wherein the probes generate a detectable fluorescent signal (creating codewords based on the fluorescence distribution within the sample); paragraphs [0008], [00141]).

-***-Continued Within the Next Supplemental Box-***-

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

-***-Continued from Previous Supplemental Box-***-

As per claim 180, California discloses the method of claim 179, further comprising exposing the secondary amplifier nucleic acids to one or more additional rounds of amplifier nucleic acids, including an ending round of amplifier nucleic acids, each able to bind the amplifier nucleic acids of the previous round (five populations of extender probes, each binding to a region of the probe population of the previous round, e.g. the second population binds a region in the first population, the third population binds a region in the second population, and so on, can be used to amplify the signal up to 5 times, with the fifth population of probes being the final population (ending round), thereby indicating exposing the tertiary probe (secondary amplifier nucleic acid) to one or more additional rounds of amplifier nucleic acids; paragraphs [0056], [00115], [00227], [00289], [00294]).

As per claim 181, California discloses the method of claim 180, and California further discloses wherein the ending amplifier nucleic acids comprise a fluorescent signaling entity (probes from all rounds, including the ending amplifier nucleic acid, are connected with a signal moiety, such as a fluorophore, that can produce a detectable signal; paragraphs [0056], [00115], [00141], [00227], [00289], [00294]).

As per claim 182, California discloses the method of claim 180, and California further discloses comprising exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity (inserting a cleavable linker between the binding sequence and fluorescent signal moiety of a probe, wherein for the next round of hybridization (ending round of amplifier nucleic acids), the visible signals are switched, thereby indicating exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity; paragraphs [0056], [00115], [00141], [00227], [00227], [00289], [00294]).

Claims 104/102-103, 109/107-108, 145/143-144, 152/150-151, 159/157-158, and 183/181-182 lack an inventive step under PCT Article 33(3) as being obvious over California in view of the publication entitled "Tandem Spinach Array for mRNA Imaging in Living Bacterial Cells" by Zhang, et al. (hereinafter 'Zhang').

As per claims 104/102-103, California discloses the method of claims 102-103, and California further discloses the ending amplifier nucleic acids linked to a fluorescent signaling entity (probes from all rounds, including the ending amplifier nucleic acid, are connected with a signal moiety, such as a fluorophore, that can produce a detectable signal; paragraphs [0056], [00115], [00141], [00227], [00289], [00294]). California does not disclose wherein the ending amplifier nucleic acids have a repeating sequence that the fluorescent signaling entity is able to bind. Zhang discloses amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind (RNA aptamer-fluorogen system, in which repeat RNA aptamers bind and induce the fluorescence of a the DFHBI fluorophore, and wherein fluorescence intensity increases with increase in aptamer repeat numbers (amplifier nucleic acids); abstract; page 2, 4th paragraph; Fig. 1C). It would have been obvious to a person of ordinary skill in the art, at the time of the invention, to have modified the method, as previously disclosed by California, for the integration of amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind, as previously disclosed by Zhang, as the previous disclosure by California includes a method comprising exposing a sample to nucleic acid probes which in turn are exposed to two or more rounds of amplifier nucleic acids, wherein the ending amplifier nucleic acids are linked to a fluorescent signaling entity, the Zhang reference discloses amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind, and this combination would provide a method of amplifying the detection signal using fewer probes with repeat sequences and hence increased fluorescence, thereby improving the efficiency of the method.

As per claims 109/107-108, California discloses the method of claims 107-108, and California further discloses the ending amplifier nucleic acids linked to a fluorescent signaling entity (probes from all rounds, including the ending amplifier nucleic acid, are connected with a signal moiety, such as a fluorophore, that can produce a detectable signal; paragraphs [0056], [00115], [00141], [00227], [00289], [00294]). California does not disclose wherein the ending amplifier nucleic acids have a repeating sequence that the fluorescent signaling entity is able to bind. Zhang discloses amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind (RNA aptamer-fluorogen system, in which repeat RNA aptamers bind and induce the fluorescence of a the DFHBI fluorophore, and wherein fluorescence intensity increases with increase in aptamer repeat numbers (amplifier nucleic acids); abstract; page 2, 4th paragraph; Fig. 1C). It would have been obvious to a person of ordinary skill in the art, at the time of the invention, to have modified the method, as previously disclosed by California, for the integration of amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind, as previously disclosed by Zhang, as the previous disclosure by California includes a method comprising exposing a sample to nucleic acid probes which in turn are exposed to two or more rounds of amplifier nucleic acids, wherein the ending amplifier nucleic acids are linked to a fluorescent signaling entity, the Zhang reference discloses amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind, and this combination would provide a method of amplifying the detection signal using fewer probes with repeat sequences and hence increased fluorescence, thereby improving the efficiency of the method.

-***-Continued Within the Next Supplemental Box-***-

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

-***-Continued from Previous Supplemental Box-***-

As per claims 145/143-144, California discloses the method of claims 143-144, and California further discloses the ending amplifier nucleic acids linked to a fluorescent signaling entity (probes from all rounds, including the ending amplifier nucleic acid, are connected with a signal moiety, such as a fluorophore, that can produce a detectable signal; paragraphs [0056], [00115], [00141], [00227], [00289], [00294]). California does not disclose wherein the ending amplifier nucleic acids have a repeating sequence that the fluorescent signaling entity is able to bind. Zhang discloses amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind (RNA aptamer-fluorogen system, in which repeat RNA aptamers bind and induce the fluorescence of a the DFHBI fluorophore, and wherein fluorescence intensity increases with increase in aptamer repeat numbers (amplifier nucleic acids); abstract; page 2, 4th paragraph; Fig. 1C). It would have been obvious to a person of ordinary skill in the art, at the time of the invention, to have modified the method, as previously disclosed by California, for the integration of amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind, as previously disclosed by Zhang, as the previous disclosure by California includes a method comprising exposing a sample to nucleic acid probes which in turn are exposed to two or more rounds of amplifier nucleic acids, wherein the ending amplifier nucleic acids are linked to a fluorescent signaling entity, the Zhang reference discloses amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind, and this combination would provide a method of amplifying the detection signal using fewer probes with repeat sequences and hence increased fluorescence, thereby improving the efficiency of the method.

As per claims 152/150-151, California discloses the method of claims 150-151, and California further discloses the ending amplifier nucleic acids linked to a fluorescent signaling entity (probes from all rounds, including the ending amplifier nucleic acid, are connected with a signal moiety, such as a fluorophore, that can produce a detectable signal; paragraphs [0056], [00115], [00141], [00227], [00289], [00294]). California does not disclose wherein the ending amplifier nucleic acids have a repeating sequence that the fluorescent signaling entity is able to bind. Zhang discloses amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind (RNA aptamer-fluorogen system, in which repeat RNA aptamers bind and induce the fluorescence of a the DFHBI fluorophore, and wherein fluorescence intensity increases with increase in aptamer repeat numbers (amplifier nucleic acids); abstract; page 2, 4th paragraph; Fig. 1C). It would have been obvious to a person of ordinary skill in the art, at the time of the invention, to have modified the method, as previously disclosed by California, for the integration of amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind, as previously disclosed by Zhang, as the previous disclosure by California includes a method comprising exposing a sample to nucleic acid probes which in turn are exposed to two or more rounds of amplifier nucleic acids, wherein the ending amplifier nucleic acids are linked to a fluorescent signaling entity, the Zhang reference discloses amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind, and this combination would provide a method of amplifying the detection signal using fewer probes with repeat sequences and hence increased fluorescence, thereby improving the efficiency of the method.

As per claims 159/157-158, California discloses the method of claims 157-158, and California further discloses the ending amplifier nucleic acids linked to a fluorescent signaling entity (probes from all rounds, including the ending amplifier nucleic acid, are connected with a signal moiety, such as a fluorophore, that can produce a detectable signal; paragraphs [0056], [00115], [00141], [00227], [00289], [00294]). California does not disclose wherein the ending amplifier nucleic acids have a repeating sequence that the fluorescent signaling entity is able to bind. Zhang discloses amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind (RNA aptamer-fluorogen system, in which repeat RNA aptamers bind and induce the fluorescence of a the DFHBI fluorophore, and wherein fluorescence intensity increases with increase in aptamer repeat numbers (amplifier nucleic acids); abstract; page 2, 4th paragraph; Fig. 1C). It would have been obvious to a person of ordinary skill in the art, at the time of the invention, to have modified the method, as previously disclosed by California, for the integration of amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind, as previously disclosed by Zhang, as the previous disclosure by California includes a method comprising exposing a sample to nucleic acid probes which in turn are exposed to two or more rounds of amplifier nucleic acids, wherein the ending amplifier nucleic acids are linked to a fluorescent signaling entity, the Zhang reference discloses amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind, and this combination would provide a method of amplifying the detection signal using fewer probes with repeat sequences and hence increased fluorescence, thereby improving the efficiency of the method.

As per claims 183/181-182, California discloses the method of claims 181-182, and California further discloses the ending amplifier nucleic acids linked to a fluorescent signaling entity (probes from all rounds, including the ending amplifier nucleic acid, are connected with a signal moiety, such as a fluorophore, that can produce a detectable signal; paragraphs [0056], [00115], [00141], [00227], [00289], [00294]). California does not disclose wherein the ending amplifier nucleic acids have a repeating sequence that the fluorescent signaling entity is able to bind. Zhang discloses amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind (RNA aptamer-fluorogen system, in which repeat RNA aptamers bind and induce the fluorescence of a the DFHBI fluorophore, and wherein fluorescence intensity increases with increase in aptamer repeat numbers (amplifier nucleic acids); abstract; page 2, 4th paragraph; Fig. 1C). It would have been obvious to a person of ordinary skill in the art, at the time of the invention, to have modified the method, as previously disclosed by California, for the integration of amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind, as previously disclosed by Zhang, as the previous disclosure by California includes a method comprising exposing a sample to nucleic acid probes which in turn are exposed to two or more rounds of amplifier nucleic acids, wherein the ending amplifier nucleic acids are linked to a fluorescent signaling entity, the Zhang reference discloses amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind, and this combination would provide a method of amplifying the detection signal using fewer probes with repeat sequences and hence increased fluorescence, thereby improving the efficiency of the method.

-***-Continued Within the Next Supplemental Box-***-

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

-***-Continued from Previous Supplemental Box-***-

Claims 110-113, 116-119, 139-140, 162-165, and 170-173 lack an inventive step under PCT Article 33(3) as being obvious over California in view of WO 2018/089438 A1 to President And Fellows of Harvard College (hereinafter 'Harvard').

As per claim 110, California discloses a method, comprising: exposing a sample to nucleic acid probes (a method of barcoding molecular targets comprising a plurality of primary probes, wherein each probe comprises a primary binding sequence that binds to a complementary target sequence (exposing) in a target nucleic acid molecule (sample); paragraphs [0008], [0056]); exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes (a plurality of bridge probes, wherein each bridge probe (primary amplifier nucleic acid) comprises a binding sequence that specifically binds to the overhang sequence of a primary probe, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0056], [00115], [00227], [00289], [00294]); exposing the primary amplifier nucleic acids to secondary amplifier nucleic acids able to bind to the primary amplifier nucleic acids (tertiary probes (secondary nucleic acids) that bind to secondary probes (primary amplifier nucleic acids) that bind to primary probes that directly bind to a target nucleic acid molecule, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0055], [0056], [0057], [00115], [00227], [00289], [00294]); determining a distribution of the nucleic acid probes within the sample using fluorescence (probes generate one or more different detectable visual fluorescent signals corresponding to the number of molecular targets (determining a distribution of probes); paragraphs [0008], [00141]); creating codewords based on the fluorescence distribution within the sample (barcoding the sample based on serial probe hybridization and visualization rounds, wherein the probes generate a detectable fluorescent signal (creating codewords based on the fluorescence distribution within the sample); paragraphs [0008], [00141]); and for at least some of the codewords, matching the codeword to a valid codeword (data from repeat rounds can validate each other, for example, a cyan data point in an image corresponding to hybridization round 2 can be matched to the image corresponding to hybridization round 3 at the same location, revealing a yellow data point, and identifying the data point as part of mRNA1 (valid codeword); paragraph [0031]) optionally wherein, if no match is found, applying error correction to the codeword to form a valid codeword (the barcoding rounds include one or more rounds of error correction, which would intrinsically be applied to the barcoding if a valid match is not found, in order to form a valid code; paragraph [0015]). California does not disclose wherein the primary amplifier nucleic acids are formed from only 3 of the 4 naturally-occurring nucleotides. Harvard discloses probes formed from only 3 of the 4 naturally-occurring nucleotides (nucleic acid probes may be made using only 3 of the 4 bases, which would eventually form naturally-occurring nucleotides; page 20, lines 1-5). It would have been obvious to a person of ordinary skill in the art, at the time of the invention, to have modified the method, as previously disclosed by California, for the integration of probes formed from only 3 of the 4 naturally-occurring nucleotides, as previously disclosed by Harvard, as the previous disclosure by California includes a method comprising sequentially exposing a sample to rounds of nucleic acid probes, wherein the second and third round probes act as primary and secondary amplifier nucleic acids, respectively, the Harvard reference discloses probes formed from only 3 of the 4 naturally-occurring nucleotides, and this combination would a method of analyzing a sample using nucleic acid probes wherein the primary amplifier probes are formed from only 3 of the 4 naturally-occurring nucleotides, thereby limiting the possibility of formation of secondary structures and ensuring uniform and faster hybridization.

As per claim 111, California and Harvard, in combination, disclose the method of claim 110, and California further discloses comprising exposing the secondary amplifier nucleic acids to one or more additional rounds of amplifier nucleic acids, including an ending round of amplifier nucleic acids, each able to bind the amplifier nucleic acids of the previous round (five populations of extender probes, each binding to a region of the probe population of the previous round, e.g. the second population binds a region in the first population, the third population binds a region in the second population, and so on, can be used to amplify the signal up to 5 times, with the fifth population of probes being the final population (ending round), thereby indicating exposing the tertiary probe (secondary amplifier nucleic acid) to one or more additional rounds of amplifier nucleic acids; paragraphs [0056], [00115], [00227], [00289], [00294]).

As per claim 112, California and Harvard, in combination, disclose the method of claim 111, and California further discloses wherein the ending amplifier nucleic acids comprise a fluorescent signaling entity (probes from all rounds, including the ending amplifier nucleic acid, are connected with a signal moiety, such as a fluorophore, that can produce a detectable signal; paragraphs [0056], [00115], [00141], [00227], [00289], [00294]).

As per claim 113, California and Harvard, in combination, disclose the method of claim 111, and California further discloses comprising exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity (inserting a cleavable linker between the binding sequence and fluorescent signal moiety of a probe, wherein for the next round of hybridization (ending round of amplifier nucleic acids), the visible signals are switched, thereby indicating exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity; paragraphs [0056], [00115], [00141], [00227], [00227], [00289], [00294]).

-***-Continued Within the Next Supplemental Box-***-

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:
-***-Continued from Previous Supplemental Box-***-

As per claim 116, California discloses a method, comprising: exposing a sample to nucleic acid probes (a method of barcoding molecular targets comprising a plurality of primary probes, wherein each probe comprises a primary binding sequence that binds to a complementary target sequence (exposing) in a target nucleic acid molecule (sample); paragraphs [0008], [0055]); exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes (a plurality of bridge probes, wherein each bridge probe (primary amplifier nucleic acid) comprises a binding sequence that specifically binds to the overhang sequence of a primary probe, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0056], [00115], [00227], [00289], [00294]); exposing the primary amplifier nucleic acids to secondary amplifier nucleic acids able to bind to the primary amplifier nucleic acids (tertiary probes (secondary nucleic acids) that bind to secondary probes (primary amplifier nucleic acids) that bind to primary probes that directly bind to a target nucleic acid molecule, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0055], [0056], [0057], [00115], [00227], [00289], [00294]); determining a distribution of the nucleic acid probes within the sample using fluorescence (probes generate one or more different detectable visual fluorescent signals corresponding to the number of molecular targets (determining a distribution of probes); paragraphs [0006], [00141]); creating codewords based on the fluorescence distribution within the sample (barcoding the sample based on serial probe hybridization and visualization rounds, wherein the probes generate a detectable fluorescent signal (creating codewords based on the fluorescence distribution within the sample); paragraphs [0008], [00141]); and for at least some of the codewords, matching the codeword to a valid codeword (data from repeat rounds can validate each other, for example, a cyan data point in an image corresponding to hybridization round 2 can be matched to the image corresponding to hybridization round 3 at the same location, revealing a yellow data point, and identifying the data point as part of mRNA1 (valid codeword); paragraph [00316]) optionally wherein, if no match is found, applying error correction to the codeword to form a valid codeword (the barcoding rounds include one or more rounds of error correction, which would intrinsically be applied to the barcoding if a valid match is not found, in order to form a valid code, paragraph [0015]). California does not disclose wherein the secondary amplifier nucleic acids are formed from only 3 of the 4 naturally-occurring nucleotides. Harverd discloses probes formed from only 3 of the 4 naturally-occurring nucleotides (nucleic acid probes may be made using only 3 of the 4 bases, which would eventually form naturally-occurring nucleotides; page 20, lines 1-5). It would have been obvious to a person of ordinary skill in the art, at the time of the invention, to have modified the method, as previously disclosed by California, for the integration of probes formed from only 3 of the 4 naturally-occurring nucleotides, as previously disclosed by Harverd, as the previous disclosure by California includes a method comprising sequentially exposing a sample to rounds of nucleic acid probes, wherein the second and third round probes act as primary and secondary amplifier nucleic acids, respectively, the Harverd reference discloses probes formed from only 3 of the 4 naturally-occurring nucleotides, and this combination would a method of analyzing a sample using nucleic acid probes wherein the secondary amplifier probes are formed from only 3 of the 4 naturally-occurring nucleotides, thereby limiting the possibility of formation of secondary structures and ensuring uniform and faster hybridization.

As per claim 117, California and Harverd, in combination, disclose the method of claim 116, further comprising exposing the secondary amplifier nucleic acids to one or more additional rounds of amplifier nucleic acids, including an ending round of amplifier nucleic acids, each able to bind the amplifier nucleic acids of the previous round (five populations of extender probes, each binding to a region of the probe population of the previous round, e.g. the second population binds a region in the first population, the third population binds a region in the second population, and so on, can be used to amplify the signal up to 5 times, with the fifth population of probes being the final population (ending round), thereby indicating exposing the tertiary probe (secondary amplifier nucleic acid) to one or more additional rounds of amplifier nucleic acids; paragraphs [0056], [00115], [00227], [00289], [00294]).

As per claim 118, California and Harverd, in combination, disclose the method of claim 117, wherein the ending amplifier nucleic acids comprise a fluorescent signaling entity (probes from all rounds, including the ending amplifier nucleic acid, are connected with a signal moiety, such as a fluorophore, that can produce a detectable signal; paragraphs [0056], [00115], [00141], [00227], [00289], [00294]).

As per claim 119, California and Harverd, in combination, disclose the method of claim 117, further comprising exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity (inserting a cleavable linker between the binding sequence and fluorescent signal moiety of a probe, wherein for the next round of hybridization (ending round of amplifier nucleic acids), the visible signals are switched, thereby indicating exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity; paragraphs [0056], [00115], [00141], [00227], [00227], [00289], [00294]).

-***-Continued Within the Next Supplemental Box-***-

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

-***-Continued from Previous Supplemental Box-***-

As per claim 139, California discloses a method, comprising: exposing a sample to nucleic acid probes (a method of barcoding molecular targets comprising a plurality of primary probes, wherein each probe comprises a primary binding sequence that binds to a complementary target sequence (exposing) in a target nucleic acid molecule (sample); paragraphs [0008], [0055]); exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes (a plurality of bridge probes, wherein each bridge probe (primary amplifier nucleic acid) comprises a binding sequence that specifically binds to the overhang sequence of a primary probe, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0056], [00115], [00227], [00269], [00294]); determining a distribution of the nucleic acid probes within the sample using fluorescence (probes generate one or more different detectable visual fluorescent signals corresponding to the number of molecular targets (determining a distribution of probes); paragraphs [0008], [00141]); and creating codewords based on the fluorescence distribution within the sample (barcoding the sample based on serial probe hybridization and visualization rounds, wherein the probes generate a detectable fluorescent signal (creating codewords based on the fluorescence distribution within the sample); paragraphs [0008], [00141]). California does not disclose wherein the primary amplifier nucleic acids are formed from only 3 of the 4 naturally-occurring nucleotides. Harvard discloses wherein the primary amplifier nucleic acids are formed from only 3 of the 4 naturally-occurring nucleotides (nucleic acid probes may be made using only 3 of the 4 bases, which would eventually form naturally-occurring nucleotides; page 20, lines 1-5). It would have been obvious to a person of ordinary skill in the art, at the time of the invention, to have modified the method, as previously disclosed by California, for the integration of probes formed from only 3 of the 4 naturally-occurring nucleotides, as previously disclosed by Harvard, as the previous disclosure by California includes a method comprising sequentially exposing a sample to rounds of nucleic acid probes, wherein the second and third round probes act as primary and secondary amplifier nucleic acids, respectively, the Harvard reference discloses probes formed from only 3 of the 4 naturally-occurring nucleotides, and this combination would a method of analyzing a sample using nucleic acid probes wherein the primary amplifier probes are formed from only 3 of the 4 naturally-occurring nucleotides, thereby limiting the possibility of formation of secondary structures and ensuring uniform and faster hybridization.

As per claim 140, California discloses a method, comprising: exposing a sample to nucleic acid probes (a method of barcoding molecular targets comprising a plurality of primary probes, wherein each probe comprises a primary binding sequence that binds to a complementary target sequence (exposing) in a target nucleic acid molecule (sample); paragraphs [0008], [0055]); exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes (a plurality of bridge probes, wherein each bridge probe (primary amplifier nucleic acid) comprises a binding sequence that specifically binds to the overhang sequence of a primary probe, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0056], [00115], [00227], [00269], [00294]); determining a distribution of the nucleic acid probes within the sample using fluorescence (probes generate one or more different detectable visual fluorescent signals corresponding to the number of molecular targets (determining a distribution of probes); paragraphs [0008], [00141]); and creating codewords based on the fluorescence distribution within the sample (barcoding the sample based on serial probe hybridization and visualization rounds, wherein the probes generate a detectable fluorescent signal (creating codewords based on the fluorescence distribution within the sample); paragraphs [0008], [00141]) and for at least some of the codewords, matching the codeword to a valid codeword (data from repeat rounds can validate each other, for example, a cyan data point in an image corresponding to hybridization round 2 can be matched to the image corresponding to hybridization round 3 at the same location, revealing a yellow data point, and identifying the data point as part of mRNA1 (valid codeword); paragraph [00316]) wherein, if no match is found, potentially applying error correction to the codeword to form a valid codeword (the barcoding rounds include one or more rounds of error correction, which would intrinsically be applied to the barcoding if a valid match is not found, in order to form a valid code; paragraph [0015]). California does not disclose wherein the primary amplifier nucleic acids are formed from only 3 of the 4 naturally-occurring nucleotides. Harvard discloses wherein the primary amplifier nucleic acids are formed from only 3 of the 4 naturally-occurring nucleotides (nucleic acid probes may be made using only 3 of the 4 bases, which would eventually form naturally-occurring nucleotides; page 20, lines 1-5). It would have been obvious to a person of ordinary skill in the art, at the time of the invention, to have modified the method, as previously disclosed by California, for the integration of probes formed from only 3 of the 4 naturally-occurring nucleotides, as previously disclosed by Harvard, as the previous disclosure by California includes a method comprising sequentially exposing a sample to rounds of nucleic acid probes, wherein the second and third round probes act as primary and secondary amplifier nucleic acids, respectively, the Harvard reference discloses probes formed from only 3 of the 4 naturally-occurring nucleotides, and this combination would a method of analyzing a sample using nucleic acid probes wherein the primary amplifier probes are formed from only 3 of the 4 naturally-occurring nucleotides, thereby limiting the possibility of formation of secondary structures and ensuring uniform and faster hybridization.

-***-Continued Within the Next Supplemental Box-***-

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:
-***-Continued from Previous Supplemental Box-***-

As per claim 162, California discloses a method, comprising: exposing a sample to nucleic acid probes (a method of barcoding molecular targets comprising a plurality of primary probes, wherein each probe comprises a primary binding sequence that binds to a complementary target sequence (exposing) in a target nucleic acid molecule (sample); paragraphs [0008], [0055]); exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes (a plurality of bridge probes, wherein each bridge probe (primary amplifier nucleic acid) comprises a binding sequence that specifically binds to the overhang sequence of a primary probe, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0056], [00115], [00227], [00289], [00294]); exposing the primary amplifier nucleic acids to secondary amplifier nucleic acids able to bind to the primary amplifier nucleic acids (tertiary probes (secondary nucleic acids) that bind to secondary probes (primary amplifier nucleic acids) that bind to primary probes that directly bind to a target nucleic acid molecule, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0055], [0056], [0057], [00115], [00227], [00289], [00294]); determining a distribution of the nucleic acid probes within the sample using fluorescence (probes generate one or more different detectable visual fluorescent signals corresponding to the number of molecular targets (determining a distribution of probes); paragraphs [0008], [00141]); and creating codewords based on the fluorescence distribution within the sample (barcoding the sample based on serial probe hybridization and visualization rounds, wherein the probes generate a detectable fluorescent signal (creating codewords based on the fluorescence distribution within the sample); paragraphs [0008], [00141]). California does not disclose wherein the primary amplifier nucleic acids are formed from only 3 of the 4 naturally-occurring nucleotides. Harvard discloses wherein the primary amplifier nucleic acids are formed from only 3 of the 4 naturally-occurring nucleotides (nucleic acid probes may be made using only 3 of the 4 bases, which would eventually form naturally-occurring nucleotides; page 20, lines 1-5). It would have been obvious to a person of ordinary skill in the art, at the time of the invention, to have modified the method, as previously disclosed by California, for the integration of probes formed from only 3 of the 4 naturally-occurring nucleotides, as previously disclosed by Harvard, as the previous disclosure by California includes a method comprising sequentially exposing a sample to rounds of nucleic acid probes, wherein the second and third round probes act as primary and secondary amplifier nucleic acids, respectively, the Harvard reference discloses probes formed from only 3 of the 4 naturally-occurring nucleotides, and this combination would a method of analyzing a sample using nucleic acid probes, wherein the primary amplifier probes are formed from only 3 of the 4 naturally-occurring nucleotides, thereby limiting the possibility of formation of secondary structures and ensuring uniform and faster hybridization.

As per claim 163, California and Harvard, in combination, disclose the method of claim 162, and California further discloses comprising: exposing the secondary amplifier nucleic acids to one or more additional rounds of amplifier nucleic acids, including an ending round of amplifier nucleic acids, each able to bind the amplifier nucleic acids of the previous round (five populations of extender probes, each binding to a region of the probe population of the previous round, e.g. the second population binds a region in the first population, the third population binds a region in the second population, and so on, can be used to amplify the signal up to 5 times, with the fifth population of probes being the final population (ending round), thereby indicating exposing the tertiary probe (secondary amplifier nucleic acid) to one or more additional rounds of amplifier nucleic acids; paragraphs [0056], [00115], [00227], [00289], [00294]).

As per claim 164, California and Harvard, in combination, disclose the method of claim 163, and California further discloses wherein the ending amplifier nucleic acids comprise a fluorescent signaling entity (probes from all rounds, including the ending amplifier nucleic acid, are connected with a signal moiety, such as a fluorophore, that can produce a detectable signal; paragraphs [0056], [00115], [00141], [00227], [00289], [00294]).

As per claim 165, California and Harvard, in combination, disclose the method of claim 163, and California further discloses comprising: exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity (inserting a cleavable linker between the binding sequence and fluorescent signal moiety of a probe, wherein for the next round of hybridization (ending round of amplifier nucleic acids), the visible signals are switched, thereby indicating exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity; paragraphs [0056], [00115], [00141], [00227], [00227], [00289], [00294]).

As per claim 170, California discloses a method, comprising: exposing a sample to nucleic acid probes; exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes (a method of barcoding molecular targets comprising a plurality of primary probes, wherein each probe comprises a primary binding sequence that binds to a complementary target sequence (exposing) in a target nucleic acid molecule (sample); paragraphs [0008], [0055]); exposing the primary amplifier nucleic acids to secondary amplifier nucleic acids able to bind to the primary amplifier nucleic acids (a plurality of bridge probes, wherein each bridge probe (primary amplifier nucleic acid) comprises a binding sequence that specifically binds to the overhang sequence of a primary probe, wherein the probes are designed to amplify the detectable signal using hybridization chain reaction (amplifier nucleic acid); paragraphs [0056], [00115], [00227], [00289], [00294]); determining a distribution of the nucleic acid probes within the sample using fluorescence (probes generate one or more different detectable visual fluorescent signals corresponding to the number of molecular targets (determining a distribution of probes); paragraphs [0008], [00141]); and creating codewords based on the fluorescence distribution within the sample (barcoding the sample based on serial probe hybridization and visualization rounds, wherein the probes generate a detectable fluorescent signal (creating codewords based on the fluorescence distribution within the sample); paragraphs [0008], [00141]). California does not disclose wherein the secondary amplifier nucleic acids are formed from only 3 of the 4 naturally-occurring nucleotides. Harvard discloses wherein the secondary amplifier nucleic acids are formed from only 3 of the 4 naturally-occurring nucleotides (nucleic acid probes may be made using only 3 of the 4 bases, which would eventually form naturally-occurring nucleotides; page 20, lines 1-5). It would have been obvious to a person of ordinary skill in the art, at the time of the invention, to have modified the method, as previously disclosed by California, for the integration of probes formed from only 3 of the 4 naturally-occurring nucleotides, as previously disclosed by Harvard, as the previous disclosure by California includes a method comprising sequentially exposing a sample to rounds of nucleic acid probes, wherein the second and third round probes act as primary and secondary amplifier nucleic acids, respectively, the Harvard reference discloses probes formed from only 3 of the 4 naturally-occurring nucleotides, and this combination would a method of analyzing a sample using nucleic acid probes wherein the secondary amplifier probes are formed from only 3 of the 4 naturally-occurring nucleotides, thereby limiting the possibility of formation of secondary structures and ensuring uniform and faster hybridization.
-***-Continued Within the Next Supplemental Box-***-

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

-***-Continued from Previous Supplemental Box-***-

As per claim 171, California and Harvard, in combination, disclose the method of claim 170, further comprising exposing the secondary amplifier nucleic acids to one or more additional rounds of amplifier nucleic acids, including an ending round of amplifier nucleic acids, each able to bind the amplifier nucleic acids of the previous round (five populations of extender probes, each binding to a region of the probe population of the previous round, e.g. the second population binds a region in the first population, the third population binds a region in the second population, and so on, can be used to amplify the signal up to 5 times, with the fifth population of probes being the final population (ending round), thereby indicating exposing the tertiary probe (secondary amplifier nucleic acid) to one or more additional rounds of amplifier nucleic acids; paragraphs [0056], [00115], [00227], [00289], [00294]).

As per claim 172, California and Harvard, in combination, disclose the method of claim 171, and California further discloses wherein the ending amplifier nucleic acids comprise a fluorescent signaling entity (probes from all rounds, including the ending amplifier nucleic acid, are connected with a signal moiety, such as a fluorophore, that can produce a detectable signal; paragraphs [0056], [00115], [00141], [00227], [00289], [00294]).

As per claim 173, California and Harvard, in combination, disclose the method of claim 171, and California further discloses comprising exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity (inserting a cleavable linker between the binding sequence and fluorescent signal moiety of a probe, wherein for the next round of hybridization (ending round of amplifier nucleic acids), the visible signals are switched, thereby indicating exposing the ending round of amplifier nucleic acids to a fluorescent signaling entity; paragraphs [0056], [00115], [00141], [00227], [00227], [00269], [00294]).

Claims 114/112-113, 120/118-119, 166/164-165, and 174/172-173 lack an inventive step under PCT Article 33(3) as being obvious over California and Harvard in view of Zhang.

As per claims 114/112-113, California and Harvard, in combination, disclose the method of any one of claims 112 or 113, but California does not disclose wherein the ending amplifier nucleic acids have a repeating sequence that the fluorescent signaling entity is able to bind. Zhang discloses amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind (RNA aptamer-fluorogen system, in which repeat RNA aptamers bind and induce the fluorescence of a the DFHBI fluorophore, and wherein fluorescence intensity increases with increase in aptamer repeat numbers (amplifier nucleic acids); abstract; page 2, 4th paragraph; Fig. 1C). It would have been obvious to a person of ordinary skill in the art, at the time of the invention, to have modified the method, as previously disclosed by California, for the integration of amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind, as previously disclosed by Zhang, as the previous disclosure by California includes a method comprising exposing a sample to nucleic acid probes which in turn are exposed to two or more rounds of amplifier nucleic acids, wherein the ending amplifier nucleic acids are linked to a fluorescent signaling entity, the Zhang reference discloses amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind, and this combination would provide a method of amplifying the detection signal using fewer probes with repeat sequences and hence increased fluorescence, thereby improving the efficiency of the method.

As per claims 120/118-119, California and Harvard, in combination, disclose the method of any one of claims 118 or 119, but California does not disclose wherein the ending amplifier nucleic acids have a repeating sequence that the fluorescent signaling entity is able to bind. Zhang discloses amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind (RNA aptamer-fluorogen system, in which repeat RNA aptamers bind and induce the fluorescence of a the DFHBI fluorophore, and wherein fluorescence intensity increases with increase in aptamer repeat numbers (amplifier nucleic acids); abstract; page 2, 4th paragraph; Fig. 1C). It would have been obvious to a person of ordinary skill in the art, at the time of the invention, to have modified the method, as previously disclosed by California, for the integration of amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind, as previously disclosed by Zhang, as the previous disclosure by California includes a method comprising exposing a sample to nucleic acid probes which in turn are exposed to two or more rounds of amplifier nucleic acids, wherein the ending amplifier nucleic acids are linked to a fluorescent signaling entity, the Zhang reference discloses amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind, and this combination would provide a method of amplifying the detection signal using fewer probes with repeat sequences and hence increased fluorescence, thereby improving the efficiency of the method.

As per claims 166/164-165, California and Harvard, in combination, disclose the method of claims 164-165, but California does not disclose wherein the ending amplifier nucleic acids have a repeating sequence that the fluorescent signaling entity is able to bind. Zhang discloses amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind (RNA aptamer-fluorogen system, in which repeat RNA aptamers bind and induce the fluorescence of a the DFHBI fluorophore, and wherein fluorescence intensity increases with increase in aptamer repeat numbers (amplifier nucleic acids); abstract; page 2, 4th paragraph; Fig. 1C). It would have been obvious to a person of ordinary skill in the art, at the time of the invention, to have modified the method, as previously disclosed by California, for the integration of amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind; as previously disclosed by Zhang, as the previous disclosure by California includes a method comprising exposing a sample to nucleic acid probes which in turn are exposed to two or more rounds of amplifier nucleic acids, wherein the ending amplifier nucleic acids are linked to a fluorescent signaling entity, the Zhang reference discloses amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind, and this combination would provide a method of amplifying the detection signal using fewer probes with repeat sequences and hence increased fluorescence, thereby improving the efficiency of the method.

-***-Continued Within the Next Supplemental Box-***-

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US19/65857

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

-***-Continued from Previous Supplemental Box-***-

As per claims 174/172-173, California and Harvard, in combination, disclose the method of claims 172-173, but California does not disclose wherein the ending amplifier nucleic acids have a repeating sequence that the fluorescent signaling entity is able to bind. Zhang discloses amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind (RNA aptamer-fluorogen system, in which repeat RNA aptamers bind and induce the fluorescence of a the DFHBI fluorophore, and wherein fluorescence intensity increases with increase in aptamer repeat numbers (amplifier nucleic acids); abstract; page 2, 4th paragraph; Fig. 1C). It would have been obvious to a person of ordinary skill in the art, at the time of the invention, to have modified the method, as previously disclosed by California, for the integration of amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind, as previously disclosed by Zhang, as the previous disclosure by California includes a method comprising exposing a sample to nucleic acid probes which in turn are exposed to two or more rounds of amplifier nucleic acids, wherein the ending amplifier nucleic acids are linked to a fluorescent signaling entity, the Zhang reference discloses amplifier nucleic acids having a repeating sequence that the fluorescent signaling entity is able to bind, and this combination would provide a method of amplifying the detection signal using fewer probes with repeat sequences and hence increased fluorescence, thereby improving the efficiency of the method.

Claims 1-2, 3/1-2, 100-103, 104/102-103, 105-108, 109/107-108, 110-113, 114/112-113, 116-119, 120/118-119, 122-133, 139-144, 145/143-144, 148-151, 152/150-151, 155-158, 159/157-158, 162-165, 166/164-165, 170-173, 174/172-173, 176-182, and 183/181-182 have industrial applicability as defined by PCT Article 33(4) because the subject matter can be made or used in industry.

Form PCT/ISA/237 (Supplemental Box) (revised January 2019)

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US19/65857

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

-***-Continued from Box No. IV: Lack of unity of invention-***-

California discloses a method, comprising: exposing a sample to nucleic acid probes (a method, comprising: exposing a sample to nucleic acid probes; paragraph [0008]); exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes (exposing the nucleic acid probes to primary amplifier nucleic acids able to bind to the nucleic acid probes; paragraphs [00115], [00125], [00227]); determining a distribution of the nucleic acid probes within the sample using fluorescence (visually identifying probes bound to targets within a sample (determining a distribution of the nucleic acid probes within the sample using fluorescence); paragraph [0008]); and creating codewords based on the fluorescence distribution within the sample (and barcoding the sample based on serial probe hybrization and visualization rounds (creating codewords based on the fluorescence distribution within the sample); paragraph [0008]); wherein the binding of primary amplifier nucleic acids and secondary amplifier nucleic acids to a target is saturatable (wherein there are N molecular targets in the sample (wherein the binding of primary amplifier nucleic acids and secondary amplifier nucleic acids to a target is saturatable); paragraphs [0008], [00115], [00227]; wherein the limited number of targets naturally creates a saturatable level of target binding sites for probes and amplifier nucleic acids); and exposing primary amplifier nucleic acids to secondary amplifier nucleic acids able to bind to the primary amplifier nucleic acids (exposing secondary probes (primary amplifier nucleic acids) to tertiary probes (secondary amplifier nucleic acids) able to bind to the secondary probes (primary amplifier nucleic acids); paragraphs [00115], [00125], [00227]).

California does not disclose: wherein the primary amplifier nucleic acids are formed from only 3 of the 4 naturally-occurring nucleotides; and wherein the secondary amplifier nucleic acids are formed from only 3 of the 4 naturally-occurring nucleotides.

Brenner discloses wherein minimally cross-hybridizing sets of oligonucleotide tags and complements thereof may be synthesized (wherein minimally cross-hybridizing sets of oligonucleotide tags and complements thereof may be synthesized; column 3, lines 36-40), depending on the size of the set desired and the degree to which cross-hybridization is sought to be minimized (depending on the size of the set desired and the degree to which cross-hybridization is sought to be minimized; column, 3, lines 36-40), using 3 of the 4 different types of nucleotides (using 3 of the 4 different types of nucleotides; column 3, lines 47-47, lines 56-59).

It would have been obvious to a person of ordinary skill in the art at the time the invention was made to have modified the disclosure of California to have used only 3 of the 4 different types of naturally occurring nucleotides for the formation of minimally cross-hybridizing sets of probes, based on the disclosure of Brenner, in order to better enable multiplexing of the detection of targets in a sample using a plurality of non-cross-hybridizing amplification probes, based on the disclosure of California, to more efficiently and effectively detect and barcode target sequences within the sample.

Since none of the special technical features of the Groups I-III inventions is found in more than one of the inventions, and since all of the shared technical features are previously disclosed by a combination of the California and Brenner references, unity of invention is lacking.

Form PCT/ISA/237 (Supplemental Box) (revised January 2019)

PCT Recordation of Search History

Case/PCT Application Number: PCT/US19/65857

CLIN Number/Technical Field of PCT Application: 0007 Biotechnology

Date(s) During Which the Search was Conducted: 2 – 16 March 2020

Date of Completion of Recordation of Search History Form: 25 March 2020

Research Analyst Initials: P1C

Search Approval Official (SAO) Initials: K1K

Field of Search/Classification Information:

IPC(8) Classifications: C12Q1/6813, 1/682, 1/6837, 1/6841; G01N 21/64, 33/533, 33/58 (2020.01)

CPC Classifications: C12Q 1/6813, 1/682, 1/6837, 1/6841; G01N 21/6486, 33/533, 33/582

Databases Searched (Patent and Non-Patent Literature (NPL), Including Sub-Databases and Files Searched) and Search Terms Used:

PatSeer (US, EP, WO, JP, DE, GB, CN, FR, KR, ES, AU, IN, CA, INPADOC Data); Google Patents; EBSCO; PubMed; Google

Terms: hybridization probe, amplifier nucleic acid, fluorescent signaling, barcode, error correction, repeat sequence

Database Search String Recordation, Including Dates of Searches:

Patent Database Search Strategy/Results:

Patseer:

ID	Search String	Records	Created Date
L21	((IC: (C12Q1/6813 OR C12Q1/682 OR C12Q1/6837 OR C12Q1/6841 OR G01N21/64 OR G01N33/533 OR G01N33/58)) OR (CPC: (C12Q1/6813 OR C12Q1/682 OR C12Q1/6837 OR C12Q1/6841 OR G01N21/6486 OR G01N33/533 OR G01N33/582))) AND (TACD: ((hybridiz*) AND (primary* OR first*) AND (secondary* OR second*) AND (probe*) AND (code*) AND (error* W5 correct*)))	875	16-Mar-20
L20	((IC: (C12Q1/6813 OR C12Q1/682 OR C12Q1/6837 OR C12Q1/6841 OR G01N21/64 OR G01N33/533 OR G01N33/58)) OR (CPC: (C12Q1/6813 OR C12Q1/682 OR C12Q1/6837 OR C12Q1/6841 OR G01N21/6486 OR G01N33/533 OR G01N33/582))) AND (TAC: ((hybridiz*) AND (primary* OR first*) AND (secondary* OR second*) AND (probe*) AND (code*) AND (error* W5 correct*)))	12	16-Mar-20
L19	((IC: (C12Q1/6813 OR C12Q1/682 OR C12Q1/6837 OR C12Q1/6841 OR G01N21/64 OR G01N33/533 OR G01N33/58)) OR (CPC: (C12Q1/6813 OR C12Q1/682 OR C12Q1/6837 OR C12Q1/6841 OR G01N21/6486 OR G01N33/533 OR G01N33/582))) AND (TACX: ((hybridiz*) AND (primary* OR first*) AND (secondary* OR second*) AND (probe*) AND (amplif*) and (code*) AND (repeat* OR repetit*) AND (error* W5 correct*)))	302	16-Mar-20
L18	((IC: (C12Q1/6813 OR C12Q1/682 OR C12Q1/6837 OR C12Q1/6841 OR G01N21/64 OR G01N33/533 OR G01N33/58)) OR (CPC: (C12Q1/6813 OR C12Q1/682 OR C12Q1/6837 OR C12Q1/6841 OR G01N21/6486 OR G01N33/533 OR G01N33/582))) AND (TAX: ((hybridiz*) AND (primary* OR first*) AND (secondary* OR second*) AND (probe*) AND (amplif*) and (code*) AND (repeat* OR repetit*) AND (error* W5 correct*)))	302	16-Mar-20
L17	((IC: (C12Q1/6813 OR C12Q1/682 OR C12Q1/6837 OR C12Q1/6841 OR G01N21/64 OR G01N33/533 OR G01N33/58)) OR (CPC: (C12Q1/6813 OR C12Q1/682 OR C12Q1/6837 OR C12Q1/6841 OR G01N21/6486 OR G01N33/533 OR G01N33/582))) AND (TACD: ((hybridiz*) AND (primary* OR first*) AND (secondary* OR second*) AND (probe*) AND (amplif*) and (code*) AND (repeat* OR repetit*) AND (error* W5 correct*)))	756	16-Mar-20
L16	((IC: (C12Q1/6813 OR C12Q1/682 OR C12Q1/6837 OR C12Q1/6841 OR G01N21/64 OR G01N33/533 OR G01N33/58)) OR (CPC: (C12Q1/6813 OR C12Q1/682 OR C12Q1/6837 OR C12Q1/6841 OR G01N21/6486 OR G01N33/533 OR G01N33/582))) AND (TAC: ((hybridiz*) AND (primary* OR first*) AND (secondary* OR second*) AND (probe*) AND (amplif*) and (code*)))	277	16-Mar-20
L15	((IC: (C12Q1/6813 OR C12Q1/682 OR C12Q1/6837 OR C12Q1/6841 OR G01N21/64 OR G01N33/533 OR G01N33/58)) OR (CPC: (C12Q1/6813 OR C12Q1/682 OR C12Q1/6837 OR C12Q1/6841 OR G01N21/6486 OR G01N33/533 OR G01N33/582))) AND (TA: ((hybridiz*) AND (primary* OR first*) AND (secondary* OR second*) AND (probe*) AND (amplif*) and (code*)))	0	16-Mar-20
L14	TACD: ((hybridiz*) AND (primary* OR first*) AND (secondary* OR second*) AND (probe*) AND (amplif*) and (code*) AND (repeat* OR repetit*) AND (error* W5 correct*))	12117	16-Mar-20
L13	TAC: ((hybridiz*) AND (primary* OR first*) AND (secondary* OR second*) AND (probe*) AND (amplif*) and (code*) AND (repeat* OR repetit*) AND (error* W5 correct*))	10	16-Mar-20

L12	TAC: ((hybridiz*) AND (primary* OR first*) AND (secondary* OR second*) AND (probe*) AND (amplif*) and (code*) AND (repeat* OR repetit*))	510	16-Mar-20
L11	TAC: ((hybridiz*) AND (primary* OR first*) AND (secondary* OR second*) AND (probe*) AND (amplif*) and (code*))	1451	16-Mar-20
L10	TA: ((hybridiz*) AND (primary* OR first*) AND (secondary* OR second*) AND (probe*) AND (amplif*) and (code*))	7	16-Mar-20
L9	TAC: (((fluorescen*_in_situ_hybrid* OR FISH) OR mer_fish) AND (primary* OR first*) AND (secondary* OR second*) AND (probe*) AND (amplif*) and (code*))	132	16-Mar-20
L8	TACD: (((multiplex*_error*_robust* W4 (fluorescen*_in_situ_hybrid* OR FISH)) OR mer_fish) AND (primary* OR first*) AND (secondary* OR second*) AND (probe*) AND (amplif*) and (code*))	134	16-Mar-20
L7	TAC: (((multiplex*_error*_robust* W4 (fluorescen*_in_situ_hybrid* OR FISH)) OR mer_fish) AND (primary* OR first*) AND (secondary* OR second*) AND (probe*) AND (amplif*) and (code*))	0	16-Mar-20
L6	(TA: ((multiplex*_error*_robust* W4 (fluorescen*_in_situ_hybrid* OR FISH)) OR mer_fish))	6	5-Mar-20
L5	(INV: ((ZHUANG W2 Xiaowei*) OR (XIA W2 Chenglong*) OR (MOFFITT W2 Jeffrey*))) AND (TACD: ((fluorescen*_in_situ_hybrid* OR FISH) AND (probe*) AND ((primary OR second*) W5 amplif*) AND (code*)))	3	5-Mar-20
L4	(ASN: (Harvard*)) AND (INV: ((ZHUANG W2 Xiaowei*) OR (XIA W2 Chenglong*) OR (MOFFITT W2 Jeffrey*))) AND (TACD: ((fluorescen*_in_situ_hybrid* OR FISH) AND (probe*) AND ((primary OR second*) W5 amplif*) AND (code*)))	3	2-Mar-20
L3	(ASN: (Harvard*)) AND (INV: ((ZHUANG W2 Xiaowei*) OR (XIA W2 Chenglong*) OR (MOFFITT W2 Jeffrey*))) AND (TAC: ((fluorescen*_in_situ_hybrid* OR FISH) AND (probe*) AND (primary_amplif*) AND (second*_amplif*) AND (code*)))	0	2-Mar-20
L2	(ASN: (Harvard*)) AND (INV: ((ZHUANG W2 Xiaowei*) OR (XIA W2 Chenglong*) OR (MOFFITT W2 Jeffrey*))) AND (TA: (fluorescen*_in_situ_hybrid* OR FISH))	1	2-Mar-20
L1	(ASN: (Harvard*)) AND (INV: ((ZHUANG W2 Xiaowei*) OR (XIA W2 Chenglong*) OR (MOFFITT W2 Jeffrey*)))	81	2-Mar-20

Google Patents: 2 March 2020

Search #	Search Terms	Results
#1	(probe) (primary) (secondary) (FISH) ((code) OR (barcode) OR (codeword)) ((fluorescence) OR (fluorescent))	8,032
#2	(probe) (primary) (secondary) (merFISH) ((code) OR (barcode) OR (codeword)) ((fluorescence) OR (fluorescent))	44
#3	(probe) (primary) (secondary) (merFISH) ((code) OR (barcode) OR (codeword)) ((fluorescence) OR (fluorescent)) (tertiary)	17

Non-Patent Literature (NPL) Search Strategy/Results:

EBSCO: 2 March 2020

Search ID#	Search Terms	Results
S6	TX hybridization AND TX probes AND TX fluorescent AND TX repeat AND TX sequence	62,062
S5	TX hybridization AND TX probes AND TX fluorescent AND TX codewords AND TX error	336
S4	TX probe AND TX FISH AND TX primary AND TX secondary AND TX barcode	7,991
S3	AU Moffitt, Jeffrey AND TX FISH	10
S2	AU Xia, Chenglong AND TX FISH	4
S1	AU Zhuang, Xiaowei AND TX FISH	20

Google: 2 – 16 March 2020

Search #	Search Terms	Results
#1	"fish" "primary probe" "secondary probe" "tertiary probe" fluorescent" "code"	5
#2	"hybridization" "primary probe" "secondary probe" "tertiary probe" fluorescence" "code"	5 results
#3	"hybridization" "amplifier" "probes" "fluorescence" "code"	80,800
#4	FISH probes with repeat sequences	6,990,000

PubMed: 2 March 2020

Search #	Search Terms	Results
#5	Search (((fluorescence[Text Word]) AND hybridization[Text Word]) AND probe[Text Word]) AND barcode[Text Word]	4
#4	Search ((FISH[Text Word]) AND probe[Text Word]) AND barcode[Text Word]	2
#3	Search (Moffitt, Jeffrey R[Author]) AND FISH[Text Word]	3
#2	Search (Xia, Chenglong[Author]) AND FISH[Text Word]	2
#1	Search (Zhuang, Xiaowei[Author]) AND FISH[Text Word]	7

Electronic Patent Application Fee Transmittal

Application Number:	15329683
Filing Date:	27-Jan-2017
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Filer:	Tani Chen/Catherine Borgnis
Attorney Docket Number:	H0498.70509US02

Filed as Small Entity

Filing Fees for U.S. National Stage under 35 USC 371

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
SUBMISSION- INFORMATION DISCLOSURE STMT	2806	1	130	130
		Total in USD ($)		**130**

Electronic Acknowledgement Receipt

EFS ID:	41716116
Application Number:	15329683
International Application Number:	
Confirmation Number:	7428
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Customer Number:	86110
Filer:	Tani Chen/Catherine Borgnis
Filer Authorized By:	Tani Chen
Attorney Docket Number:	H0498.70509US02
Receipt Date:	22-JAN-2021
Filing Date:	27-JAN-2017
Time Stamp:	11:55:42
Application Type:	U.S. National Stage under 35 USC 371

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$130
RAM confirmation Number	E20211LB56199729
Deposit Account	232825
Authorized User	Catherine Borgnis
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows: 37 CFR 1.17 (Patent application and reexamination processing fees)	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	H049870509US02-TRN-TC.pdf	26195 6231c08d0f3885cfe7b3a89ad70de118c90cb70e	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	H049870509US02-FEE-TC.pdf	32832 329811b1d56f8030572fcb6b6ce35cea925fdf74	no	1
Warnings:					
Information:					
3	Transmittal Letter	H049870509US02-IDSBODY-TC.pdf	35560 3f5543c7ad89228573120a40513bd9c9490d8419	no	4
Warnings:					
Information:					
4	Information Disclosure Statement (IDS) Form (SB08)	H049870509US02-IDS1449-TC.pdf	34803 da590e8c1aad120e491d1329d3bb0dede774b484	no	2
Warnings:					
Information:					
This is not an USPTO supplied IDS fillable form					
5	Foreign Reference	WO2018089438A1.pdf	5996719 b163f77d0bea04d3393265a4f464bdeb5dd31624	no	71
Warnings:					
Information:					
6	Other Reference-Patent/App/Search documents	H049870509EP00_EP_OA_dtd_10-16-2020.pdf	94747 6e682b4ed07cb46df744631cdd566e70f97c877f	no	4
Warnings:					

Information:					
7	Other Reference-Patent/App/Search documents	H049870510CN00_CN_OA_mailed_08-19-2020.pdf	425215 0813ca1345d9b6977a778529b1552a70fbe68dce	no	11
Warnings:					
Information:					
8	Other Reference-Patent/App/Search documents	H049870510EP00_EP_OA_mailed_08-19-2020.pdf	120668 d31241df7a2afd4d1e9bdb4f010ba57efd44f39f	no	4
Warnings:					
Information:					
9	Other Reference-Patent/App/Search documents	H049870630EP00_EESR_mailed_05-20-20.pdf	313836 2034cfecbb6a8c03bf92e0672bb5d6d2f2afaa35	no	9
Warnings:					
Information:					
10	Other Reference-Patent/App/Search documents	H049870630EP00_Supp_EP_SR_mailed_06-09-2020.pdf	47227 f63cf51d8b5eb3a2ac395eb74f3039cfae781e28	no	1
Warnings:					
Information:					
11	Other Reference-Patent/App/Search documents	H049870645EP00_EESR_mailed_07-20-2020.pdf	548851 b55e8e3749042eb5a3af3fb9b61bbaee342d8c13	no	14
Warnings:					
Information:					
12	Other Reference-Patent/App/Search documents	H049870645EP00_EP_OA_mailed_08-06-2020.pdf	47229 402a394857dc734eab4b678d32109fd7db68d194	no	1
Warnings:					
Information:					
13	Other Reference-Patent/App/Search documents	H049870693WO00_ISR_mailed_07-17-2020.pdf	1537707 e128fd2f98b008238e518bea2550d5c7eb97e04b	no	17
Warnings:					
Information:					

14	Non Patent Literature	CHEN_et_al_Science_2015_v348_p6090.pdf	3368166 5eb22b049020c109672e239b36006fe071b13bfe	no	16
Warnings:					
Information:					
15	Non Patent Literature	Chen_et_al_Science_Supp_Material_2015_v347_p1.pdf	1312392 72af14d44470c157a4c5e0eda8952eb541ec52b5	no	18
Warnings:					
Information:					
16	Non Patent Literature	CHUNG_et_al_Nat_Meth_2013_v10_p508.pdf	1398761 c75dd61d5a3d9d78dba8569a6ca31b79c33437a8	no	7
Warnings:					
Information:					
17	Non Patent Literature	TSANOV_et_al_Nuc_Acid_Res_2016_v44_pe165.pdf	2126829 156692bd9a2f7d0d1872f7aed9798a6590c19534	no	11
Warnings:					
Information:					
18	Non Patent Literature	YANG_et_al_Cell_2014_v158_p945.pdf	10968823 c2588814bf68e330b293e1afa6bc119e41f3d47e	no	27
Warnings:					
Information:					
19	Non Patent Literature	ZHANG_et_al_Sci_Rep_2015_v5_p17295.pdf	3161086 055cbb2f17e79db98deb425300bb0f8ce5231508	no	9
Warnings:					
Information:					
20	Other Reference-Patent/App/Search documents	H049870680WO00_ISR_mailed_05-01-2020.pdf	4016803 2d6788100a2c31188aa36c24eba318ecade5d1d0	no	32
Warnings:					
Information:					

21	Fee Worksheet (SB06)	fee-info.pdf	30599 936f9a06fdbdcb1e1e7af4e8f6ad1ae556160860	no	2
Warnings:					
Information:					
			Total Files Size (in bytes):	35645048	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.
National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.
New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

DOCKET NO.: H0498.70509US02

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	Xiaowei Zhuang
Application No.:	15/329,683
Confirmation No.:	7428
Filed:	January 27, 2017
For:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
Examiner:	F. W. Lu
Art Unit:	1634

Certificate of Electronic Filing under 37 C.F.R. § 1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: January 22, 2021 Electronic Signature for Catherine E. Borgnis: /Catherine E. Borgnis/

MAIL STOP AMENDMENT
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

STATEMENT FILED PURSUANT TO THE DUTY OF DISCLOSURE UNDER 37 C.F.R. §§ 1.56, 1.97 AND 1.98

Sir:

Pursuant to the duty of disclosure under 37 C.F.R. §§ 1.56, 1.97 and 1.98, the undersigned requests consideration of this Information Disclosure Statement.

PART I: Compliance with 37 C.F.R. § 1.97

This Information Disclosure Statement has been filed more than three months after the filing date of this application and after the mailing date of a first Office Action, but before the mailing date of any of a Final Action under 37 C.F.R. § 1.113, a Notice of Allowance under 37 C.F.R. § 1.311, or an action that otherwise closes prosecution in this application.

Please charge our Credit Card in the amount of $130.00 covering the fee set forth in 37 C.F.R. § 1.17(p).

PART II: Information Cited

The undersigned hereby makes of record in the above-identified application the information listed on the attached form PTO-1449 (modified PTO/SB/08). The order of presentation of the references should not be construed as an indication of the importance of the references.

The undersigned hereby makes the following additional information of record in the above-identified application.

The undersigned would like to bring to the Examiner's attention the following co-pending applications that may contain subject matter related to this application:

Serial No.	Filing Date	Inventor(s)	Docket No.
*16/253,919	1/22/2019	Zhuang et al.	H0498.70510US02
*16/347,874	5/7/2019	Zhuang et al.	H0498.70630US01
*16/616,833	11/25/2019	Zhuang et al.	H0498.70637US01
*16/348,071	5/7/2019	Zhuang et al.	H0498.70645US01

*A copy of this reference is not provided as the Office has waived the requirement under 37 C.F.R. § 1.98(a)(2)(iii) for submitting a copy of a cited U.S. patent application if it is scanned to the Image File Wrapper system and is available on Private PAIR .

The undersigned would like to bring to the Examiner's attention the enclosed search report or other communication from a corresponding or related International or Foreign National Application:

Serial No.	Date of Mailing	Type(s) of Communication	Docket No.
EP 15827358.1	**10-16-2020**	**European Office Action**	**H0498.70509EP00**
CN 201580051029.5	**08-19-2020**	**Chinese Office Action**	**H0498.70510CN00**
EP 15828133.7	**08-19-2020**	**European Office Action**	**H0498.70510EP00**
EP 17869122.6	**05-20-2020**	**Extended European Search Report**	**H0498.70630EP00**
EP 17869122.6	**06-09-2020**	**Supplementary European Search Report**	**H0498.70630EP00**
EP 17869515.1	**07-20-2020**	**Extended European Search Report**	**H0498.70645EP00**
EP 17869515.1	**08-06-2020**	**European Office Action**	**H0498.70645EP00**
PCT/US2019/065857	**05-01-2020**	**International Search Report and Written Opinion**	**H0498.70680WO00**
PCT/US20/28632	**07-16-2020**	**International Search Report and Written Opinion**	**H0498.70693WO00**

PART III: Remarks

Documents cited anywhere in the Information Disclosure Statement are enclosed unless otherwise indicated. It is respectfully requested that:

1. The Examiner consider completely the cited information, along with any other information, in reaching a determination concerning the patentability of the present claims;

2. The enclosed form PTO-1449 (modified PTO/SB/08) be signed by the Examiner to evidence that the cited information has been fully considered by the United States Patent and Trademark Office during the examination of this application;

3. The citations for the information be printed on any patent which issues from this application.

By submitting this Information Disclosure Statement, the undersigned makes no representation that a search has been performed, of the extent of any search performed, or that more relevant information does not exist.

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, material to patentability as defined in 37 C.F.R. § 1.56(b).

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, in fact, prior art as defined by 35 U.S.C. § 102.

Notwithstanding any statements by the undersigned, the Examiner is urged to form his or her own conclusion regarding the relevance of the cited information.

An early and favorable action is hereby requested.

The Director is hereby authorized to charge any deficiency or credit any overpayment in the fees occasioned by the filing of this Information Disclosure Statement to our Deposit Account No. 23/2825 under Docket No. H0498.70509US02 from which the undersigned is authorized to draw.

Respectfully submitted,

Electronic signature: /Tani Chen/
Tani Chen, Sc.D.
Registration No.: 52,728
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

Docket No.: H0498.70509US02
Date: January 22, 2021

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/329,683-Conf. #7428
Filing Date	January 27, 2017
First Named Inventor	Xiaowei Zhuang
Art Unit	1634
Examiner Name	F. W. Lu
Attorney Docket Number	H0498.70509US02

ENCLOSURES (*Check all that apply*)

- [x] Fee Transmittal Form
  - [ ] Fee Attached
- [ ] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [x] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) (please Identify below):

  Form PTO-1449
  Copies of Cited References

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Tani Chen/		
Printed name	Tani Chen, Sc.D.		
Date	January 22, 2021	Reg. No.	52,728

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: January 22, 2021 Electronic Signature for Catherine E. Borgnis: /Catherine E. Borgnis/

8892646.1

FEE TRANSMITTAL

Complete if known	
Application Number	15/329,683-Conf. #7428
Filing Date	January 27, 2017
First Named Inventor	Xiaowei Zhuang
Examiner Name	F. W. Lu
Art Unit	1634
Practitioner Docket No.	H0498.70509US02

[x] Applicant asserts small entity status. See 37 CFR 1.27.

[ ] Applicant certifies micro entity status. See 37 CFR 1.29. Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

TOTAL AMOUNT OF PAYMENT ($)130.00

METHOD OF PAYMENT (check all that apply)

[ ] Check [x] Credit Card [ ] Money Order [ ] None [ ] Other (please identify): _____

[x] Deposit Account Deposit Account Number: 23/2825 Deposit Account Name: Wolf, Greenfield & Sacks, P.C.

For the above-identified deposit account, the Director is hereby authorized to (check all that apply):

[x] Charge fee(s) indicated below

[ ] Charge fee(s) indicated below, **except for the filing fee**

[x] Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17, **except for any excess claims fees or multiple dependent claim fee**

[ ] Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17

[x] Credit any overpayment of fee(s)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES (U = undiscounted fee; S = small entity fee; M = micro entity fee)

	FILING FEES			SEARCH FEES			EXAMINATION FEES			
Application Type	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**Fees Paid ($)**
Utility	320	160*	80	700	350	175	800	400	200	
Design	220	110	55	160	80	40	640	320	160	
Plant	220	110	55	440	220	110	660	330	165	
Reissue	320	160	80	700	350	175	2,320	1,160	580	
Provisional	300	150	75	0	0	0	0	0	0	

* The $160 small entity status filing fee for a utility application is further reduced to $80 for a small entity status applicant who files the application via EFS-Web.

2. EXCESS CLAIM FEES

Fee Description	**Undiscounted Fee ($)**	**Small Entity Fee ($)**	**Micro Entity Fee ($)**
Each claim over 20 (including Reissues)	100	50	25
Each independent claim over 3 (including Reissues)	480	240	120
Multiple dependent claims	860	430	215

Total Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**
_____ - 20 or HP =	_____ x	_____ =	_____

HP = highest number of total claims paid for, if greater than 20.

Indep. Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**
_____ - 3 or HP =	_____ x	_____ =	_____

HP = highest number of independent claims paid for, if greater than 3.

Multiple Dependent Claims	
Fee ($)	**Fee Paid ($)**
_____	_____

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper (excluding electronically filed sequence or computer listings under 37 CFR 1.52(e)), the application size fee due is $420 ($210 for small entity) ($105 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets	**Extra Sheets**	**Number of each additional 50 or fraction thereof**	**Fee ($)**	**Fee Paid ($)**
_____ - 100 =	_____ /50 =	_____ (round **up** to a whole number) x	_____ =	_____

4. OTHER FEE(S)

	Fees Paid ($)
Non-English specification, $130 fee (no small or micro entity discount)	
Non-electronic filing fee under 37 CFR 1.16(t) for a utility application, $400 fee ($200 small or micro entity)	
Other (e.g., late filing surcharge): 2806 Submission of an Information Disclosure Statement	130.00

SUBMITTED BY

Signature	/Tani Chen/	Registration No. (Attorney/Agent)	52,728	Telephone	617.646.8000
Name (Print/Type)	Tani Chen, Sc.D.			Date	January 22, 2021

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: January 22, 2021 Electronic Signature for Catherine E. Borgnis: /Catherine E. Borgnis/

8892645.1

PTO/SB/06 (09-11)
Approved for use through 1/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PATENT APPLICATION FEE DETERMINATION RECORD Substitute for Form PTO-875	Application or Docket Number 15/329,683	Filing Date 01/27/2017	☐ To be Mailed

ENTITY: ☑ LARGE ☐ SMALL ☐ MICRO

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA		RATE ($)	FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A		N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A		N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 =	*		x $80 =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 =	*		x $420 =	
☐ APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).				
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))					
* If the difference in column 1 is less than zero, enter "0" in column 2.				TOTAL	

APPLICATION AS AMENDED - PART II

		(Column 1)		(Column 2)	(Column 3)			
AMENDMENT	01/22/2021	CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	* 16	Minus	** 20	= 0		x $100 =	0
	Independent (37 CFR 1.16(h))	* 1	Minus	*** 3	= 0		x $480 =	0
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	0

		(Column 1)		(Column 2)	(Column 3)			
AMENDMENT		CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	*	Minus	**	=		x $0 =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=		x $0 =	
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.	LIE
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".	/MARISSA R BLYTHER/
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".	
The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.	

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/329,683	01/27/2017	Xiaowei Zhuang	H0498.70509US02	7428

86110 7590 02/24/2021

Wolf, Greenfield & Sacks, P.C. (Harvard)
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER	
LU, FRANK WEI MIN	

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
02/24/2021	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

H0498_eOfficeAction@WolfGreenfield.com
Patents_eOfficeAction@WolfGreenfield.com

Applicant-Initiated Interview Summary	**Application No.** 15/329,683	**Applicant(s)** Zhuang et al.		
	Examiner FRANK W LU	**Art Unit** 1634	**AIA (First Inventor to File) Status** Yes	**Page** 1 of 1

All Participants (applicant, applicants representative, PTO personnel)	**Title**	**Type**
FRANK W LU	Primary Examiner	Telephonic
Dr. Jeffrey moffit	Inventor	
Tani Chen	Attorney of Record	

Date of Interview: 18 February 2021

Issues Discussed:

Proposed Amendment(s)

Dr. moffit, Mr. Chen, and the examiner discussed the inventions and possible amendment. The interview helped the examiner to better understand the inventions of applicants.

/FRANK W LU/ Primary Examiner, Art Unit 1634	February 18, 2021

Applicant is reminded that a complete written statement as to the substance of the interview must be made of record in the application file. It is the applicants responsibility to provide the written statement, unless the interview was initiated by the Examiner and the Examiner has indicated that a written summary will be provided. See MPEP 713.04
Please further see:
MPEP 713.04
Title 37 Code of Federal Regulations (CFR) § 1.133 Interviews, paragraph (b)
37 CFR § 1.2 Business to be transacted in writing

Applicant recordation instructions: The formal written reply to the last Office action must include the substance of the interview. (See MPEP section 713.04). If a reply to the last Office action has already been filed, applicant is given a non-extendable period of the longer of one month or thirty days from this interview date, or the mailing date of this interview summary form, whichever is later, to file a statement of the substance of the interview.

Examiner recordation instructions: Examiners must summarize the substance of any interview of record. A complete and proper recordation of the substance of an interview should include the items listed in MPEP 713.04 for complete and proper recordation including the identification of the general thrust of each argument or issue discussed, a general indication of any other pertinent matters discussed regarding patentability and the general results or outcome of the interview, to include an indication as to whether or not agreement was reached on the issues raised.

Docket No.: H0498.70509US02
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	Xiaowei Zhuang
Application No.:	15/329,683
Confirmation No.:	7428
Filed:	January 27, 2017
For:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
Examiner:	F. W. Lu
Art Unit:	1634

Certificate of Electronic Filing under 37 C.F.R. § 1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: March 26, 2021 Electronic Signature for Angela M. Griffith: /Angela M. Griffith/

SUPPLEMENTAL AMENDMENT

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

INTRODUCTORY COMMENTS

Further to the Response filed January 22, 2021, please amend the above-identified U.S. patent application as follows. This Supplemental Amendment is being submitted pursuant to the Examiner's express request made to the undersigned on March 25, 2021 by telephone interview.

The Claims are reflected in the listing of claims which begins on page 2 of this paper.

Remarks begin on page 6 of this paper.

IN THE CLAIMS

Please replace all prior versions, and listings, of claims in the application with the following list of claims. Additions are indicated by underlining and deletions are indicated by strikeouts and/or double bracketing.

1. (Previously Presented) A method for detecting nucleic acid targets, comprising:

(a) exposing a sample suspected of containing at least some of a plurality of nucleic acid targets to a plurality of encoding nucleic acid probes,

wherein at least some of the plurality of encoding nucleic acid probes comprise a target sequence that specifically hybridizes to a nucleic acid target in the plurality of nucleic acid targets and one or more read sequences,

wherein unique combinations of one or more read sequences correspond to valid codewords, each of the valid codewords associated to each of the plurality of nucleic acid targets, wherein the valid codeword consists of a series of positions and values for each position, and

each read sequence is associated with each of the values, or a subset thereof, at each of the positions of the valid codewords;

(b) exposing the sample to a first round of one or more readout nucleic acid probes, wherein at least some of the readout nucleic acid probes comprise a first portion and a second portion, the first portion comprising a sequence that specifically hybridizes to one of the read sequences, and the second portion comprising a fluorescent signaling entity;

(c) identifying nucleic acid targets in the sample, comprising:

detecting fluorescent signals emitted from the fluorescent signaling entities of bound readout nucleic acid probes in at least some locations within the sample to determine a codeword in at least some locations within the sample based on the read sequences present on the bound encoding nucleic acid probes;

identifying the codeword at each location by comparing the codeword to the valid codewords associated with the plurality of nucleic acid targets, and if no match is found, applying error correction to the codeword to match one of the valid codewords or discarding

the codeword; and

assigning the identity of some of the plurality of nucleic acid targets in the sample based on the identity of the nucleic acid targets associated with the valid codewords to which the identified codewords were matched.

2-33. (Cancelled)

34. (Currently Amended) The method of claim 1, further comprising inactivating at least some of the fluorescent signaling entity and/or removing at least some of the fluorescent signaling entity from the sample after said exposing the sample to the one or more readout nucleic acid probes and ~~imaging the sample to determine the fluorescence signal of~~ detecting fluorescent signals emitted from the fluorescent signaling entities.

35-70. (Cancelled)

71. (Previously presented) The method of claim 1, comprising determining the fluorescent signaling entity at a resolution better than 300 nm.

72-74. (Cancelled)

75. (Previously presented) The method of claim 1, wherein the sample comprises a cell.

76. (Cancelled)

77. (Previously presented) The method of claim 75, wherein the cell is fixed.

78. (Previously presented) The method of claim 1, comprising determining the fluorescent signal entity by imaging at least a portion of the sample using a light microscope.

79-176. (Cancelled)

177. (Currently Amended) The method of claim 1, further comprising:

exposing the sample to a second round of one or more readout nucleic acid probes, wherein at least some of the readout nucleic acid probes of the second round comprise a first portion and a second portion, the first portion comprising a sequence that specifically hybridizes to one of the read sequences different from the read sequences of the first round, and the second portion comprising a fluorescent signaling entity; and

~~imaging the sample to determine a fluorescence signal of~~ detecting fluorescent signals emitted from the fluorescent signaling entity of the second round of one or more readout nucleic acid probes in at least some locations in the sample.

178. (Previously Presented) The method of claim 1, wherein the values at each position in the codewords are binary values.

179. (Previously Presented) The method of claim 1, further comprising determining an abundance and/or spatial distribution of the nucleic acid targets based on the codewords matched to the valid codewords in each location within the sample.

180. (Currently Amended) The method of claim 1, further comprising removing unbound readout nucleic acid probes prior to said ~~imaging the sample~~ detecting fluorescent signals.

181. (Currently Amended) The method of claim 34, wherein said exposing the sample to one or more readout nucleic acid probes, said ~~imaging the sample~~ detecting fluorescent signals, and said inactivating at least some of the fluorescent signaling entities on the read out probes are performed to assign each of the values, or subset thereof, at each of the positions within the codewords.

182. (Cancelled)

183. (Previously Presented) The method of claim 1, wherein the valid codewords form an error-checking or an error-correcting code space.

184. (Previously Presented) The method of claim 1, wherein in comparing the codeword to the valid codewords in step (c), no match is found when the codeword differs from the valid codeword compared by one or more errors.

185. (Previously Presented) The method of claim 1, wherein in comparing the codeword to the valid codewords in step (c), no match is found when the codeword differs from the valid codeword compared by one error.

186. (Previously Presented) The method of claim 1, wherein in comparing the codeword to the valid codewords in step (c), no match is found when the codeword differs from the valid codeword compared by two errors.

187. (Previously Presented) The method of claim 1, wherein comparing the codeword to the valid codewords in step (c) comprises the values at each of the positions of the codewords.

REMARKS

In response to the Examiner's request of March 25, 2021 by telephone interview to the undersigned, this Supplemental Amendment is filed to amend several claims, as follows. In particular, in claim 34, "imaging the sample to determine the fluorescence signal of" has been amended to "detecting fluorescent signals." Similar amendments have been made to claim 177. In addition, in claim 180, "imaging the sample" has been amended to "detecting fluorescent signals." Similar amendments have been made to claim 181.

No new matter has been added by way of these amendments. After these amendments, claims 1, 34, 71, 75, 77, 78, and 177-182 remain pending for examination.

Favorable action is respectfully requested. If, for any reason, the Examiner is of the opinion that a telephone conversation with the undersigned would expedite prosecution, the Examiner is kindly invited to contact the undersigned at 617-646-8000.

If there is a fee occasioned by this Supplemental Amendment, including an extension fee, the Director is hereby authorized to charge any deficiency or credit any overpayment in the fees filed, asserted to be filed or which should have been filed herewith to Deposit Account No. 23/2825, under Docket No. H0498.70509US02, from which the undersigned is authorized to draw.

Dated: March 26, 2021

Respectfully submitted,

Electronic signature: /Tani Chen/
Tani Chen, Sc.D.
Registration No.: 52,728
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

Electronic Acknowledgement Receipt

EFS ID:	42289628
Application Number:	15329683
International Application Number:	
Confirmation Number:	7428
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Customer Number:	86110
Filer:	Tani Chen/Angela Griffith
Filer Authorized By:	Tani Chen
Attorney Docket Number:	H0498.70509US02
Receipt Date:	26-MAR-2021
Filing Date:	27-JAN-2017
Time Stamp:	09:51:11
Application Type:	U.S. National Stage under 35 USC 371

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Transmittal Letter	H049870509US02-TRN-TC.pdf	26136 8e3a55bca806b663217c23183c45ed8b7d647d60	no	1
Warnings:					

Information:					
2		H049870509US02-AMN-TC.pdf	37280 9ffcc19e2ab66cbd1d521b3f4347f4893423 1ce7	yes	6

Multipart Description/PDF files in .zip description		
Document Description	**Start**	**End**
Supplemental Response or Supplemental Amendment	1	1
Claims	2	5
Applicant Arguments/Remarks Made in an Amendment	6	6

Warnings:	
Information:	
Total Files Size (in bytes):	63416

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/329,683-Conf. #7428
Filing Date	January 27, 2017
First Named Inventor	Xiaowei Zhuang
Art Unit	1634
Examiner Name	F. W. Lu
Attorney Docket Number	H0498.70509US02

ENCLOSURES (*Check all that apply*)

- [ ] Fee Transmittal Form
  - [ ] Fee Attached
- [x] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [ ] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [ ] Other Enclosure(s) (please Identify below):

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Tani Chen/		
Printed name	Tani Chen, Sc.D.		
Date	March 26, 2021	Reg. No.	52,728

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: March 26, 2021 Electronic Signature for Angela M. Griffith: /Angela M. Griffith/

9078508.1

PTO/SB/06 (09-11)
Approved for use through 1/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PATENT APPLICATION FEE DETERMINATION RECORD Substitute for Form PTO-875	Application or Docket Number 15/329,683	Filing Date 01/27/2017	☑ To be Mailed

ENTITY: ☐ LARGE ☑ SMALL ☐ MICRO

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA		RATE ($)	FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A		N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A		N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 =	*		x $40 =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 =	*		x $210 =	
☐ APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).				
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))					
* If the difference in column 1 is less than zero, enter "0" in column 2.				TOTAL	

APPLICATION AS AMENDED - PART II

		(Column 1)		(Column 2)	(Column 3)			
AMENDMENT	03/26/2021	CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	* 16	Minus	** 20	= 0		x $50 =	0
	Independent (37 CFR 1.16(h))	* 1	Minus	*** 3	= 0		x $240 =	0
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	0

		(Column 1)		(Column 2)	(Column 3)			
AMENDMENT		CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	*	Minus	**	=		x $0 =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=		x $0 =	
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.	LIE
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".	/ANGELA D JOHNSON/
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".	
The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.	

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/329,683	01/27/2017	Xiaowei Zhuang	H0498.70509US02	7428

86110 7590 04/26/2021

Wolf, Greenfield & Sacks, P.C. (Harvard)
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER	
LU, FRANK WEI MIN	

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
04/26/2021	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

H0498_eOfficeAction@WolfGreenfield.com
Patents_eOfficeAction@WolfGreenfield.com

Applicant-Initiated Interview Summary	Application No. 15/329,683	Applicant(s) Zhuang et al.		
	Examiner FRANK W LU	Art Unit 1634	AIA (First Inventor to File) Status Yes	Page 1 of 1

All Participants (applicant, applicants representative, PTO personnel)	Title	Type
FRANK W LU	Primary Examiner	Telephonic
Dr.Tani Chen (Reg. No. 52,728)	Attorney of Record	
Ho Jung Cho (Reg. No. 72,755)	McNeill Baur PLLC	

Date of Interview: 19 April 2021

Issues Discussed:

Proposed Amendment(s)

The examiner and applicant's representatives discussed possible amendment in claim 1. Applicant will send another amendment on claim 1 to the examiner for review.

/FRANK W LU/ Primary Examiner, Art Unit 1634	April 19, 2021

Applicant is reminded that a complete written statement as to the substance of the interview must be made of record in the application file. It is the applicants responsibility to provide the written statement, unless the interview was initiated by the Examiner and the Examiner has indicated that a written summary will be provided. See MPEP 713.04
Please further see:
MPEP 713.04
Title 37 Code of Federal Regulations (CFR) § 1.133 Interviews, paragraph (b)
37 CFR § 1.2 Business to be transacted in writing

Applicant recordation instructions: The formal written reply to the last Office action must include the substance of the interview. (See MPEP section 713.04). If a reply to the last Office action has already been filed, applicant is given a non-extendable period of the longer of one month or thirty days from this interview date, or the mailing date of this interview summary form, whichever is later, to file a statement of the substance of the interview.

Examiner recordation instructions: Examiners must summarize the substance of any interview of record. A complete and proper recordation of the substance of an interview should include the items listed in MPEP 713.04 for complete and proper recordation including the identification of the general thrust of each argument or issue discussed, a general indication of any other pertinent matters discussed regarding patentability and the general results or outcome of the interview, to include an indication as to whether or not agreement was reached on the issues raised.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/329,683-Conf. #7428
Filing Date	January 27, 2017
First Named Inventor	Xiaowei Zhuang
Art Unit	1634
Examiner Name	F. W. Lu
Attorney Docket Number	H0498.70509US02

ENCLOSURES (*Check all that apply*)

- [ ] Fee Transmittal Form
  - [ ] Fee Attached
- [ ] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [ ] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) (please Identify below):

  Response to Interview Summary

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Tani Chen/		
Printed name	Tani Chen, Sc.D.		
Date	May 20, 2021	Reg. No.	52,728

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: May 20, 2021 Electronic Signature for: /Eileen M. MacKenzie/

9254727.1

Docket No.: H0498.70509US02
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	Xiaowei Zhuang
Application No.:	15/329,683
Confirmation No.:	7428
Filed:	January 27, 2017
For:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
Examiner:	F. W. Lu
Art Unit:	1634

Certificate of Electronic Filing under 37 C.F.R. § 1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: ____May 20, 2021_ Electronic Signature for: /Eileen M. MacKenzie/

RESPONSE TO INTERVIEW SUMMARY

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

This statement is hereby filed in order to comply with M.P.E.P. § 713.04 in response to the Interview Summary dated April 26, 2021. The undersigned agrees that the Interview Summary reflects the substance of the interview.

In addition, it should be noted that while the undersigned represents the Applicant, Dr. Ho Jung Cho represents a licensee, rather than the Applicant.

If, for any reason, the Examiner is of the opinion that a telephone conversation with the undersigned would expedite prosecution, the Examiner is kindly invited to contact the undersigned at 617-646-8000.

Dated: May 20, 2021

Respectfully submitted,

Electronic signature: /Tani Chen/
Tani Chen, Sc.D.
Registration No.: 52,728
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

Electronic Acknowledgement Receipt

EFS ID:	42781283
Application Number:	15329683
International Application Number:	
Confirmation Number:	7428
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Customer Number:	86110
Filer:	Tani Chen/Eileen MacKenzie
Filer Authorized By:	Tani Chen
Attorney Docket Number:	H0498.70509US02
Receipt Date:	20-MAY-2021
Filing Date:	27-JAN-2017
Time Stamp:	19:30:29
Application Type:	U.S. National Stage under 35 USC 371

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	H049870509US02-TRN-TC.pdf	26252 e35458bb763bacfa98322ac936c969b235fa8118	no	1
Warnings:					

Information:					
2	Applicant summary of interview with examiner	H049870509US02-ISUM-TC.pdf	26664 0b4fd561403d297c851bf88eb75fac6a7817db3c	no	2
Warnings:					
Information:					
			Total Files Size (in bytes):	52916	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

<u>New Applications Under 35 U.S.C. 111</u>
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.
<u>National Stage of an International Application under 35 U.S.C. 371</u>
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.
<u>New International Application Filed with the USPTO as a Receiving Office</u>
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

NOTICE OF ALLOWANCE AND FEE(S) DUE

86110 7590 05/26/2021
Wolf, Greenfield & Sacks, P.C. (Harvard)
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER
LU, FRANK WEI MIN

ART UNIT	PAPER NUMBER
1634	

DATE MAILED: 05/26/2021

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/329,683	01/27/2017	Xiaowei Zhuang	H0498.70509US02	7428

TITLE OF INVENTION: SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	SMALL	$600	$0.00	$0.00	$600	08/26/2021

THE APPLICATION IDENTIFIED ABOVE HAS BEEN EXAMINED AND IS ALLOWED FOR ISSUANCE AS A PATENT. PROSECUTION ON THE MERITS IS CLOSED. THIS NOTICE OF ALLOWANCE IS NOT A GRANT OF PATENT RIGHTS. THIS APPLICATION IS SUBJECT TO WITHDRAWAL FROM ISSUE AT THE INITIATIVE OF THE OFFICE OR UPON PETITION BY THE APPLICANT. SEE 37 CFR 1.313 AND MPEP 1308.

THE ISSUE FEE AND PUBLICATION FEE (IF REQUIRED) MUST BE PAID WITHIN THREE MONTHS FROM THE MAILING DATE OF THIS NOTICE OR THIS APPLICATION SHALL BE REGARDED AS ABANDONED. THIS STATUTORY PERIOD CANNOT BE EXTENDED. SEE 35 U.S.C. 151. THE ISSUE FEE DUE INDICATED ABOVE DOES NOT REFLECT A CREDIT FOR ANY PREVIOUSLY PAID ISSUE FEE IN THIS APPLICATION. IF AN ISSUE FEE HAS PREVIOUSLY BEEN PAID IN THIS APPLICATION (AS SHOWN ABOVE), THE RETURN OF PART B OF THIS FORM WILL BE CONSIDERED A REQUEST TO REAPPLY THE PREVIOUSLY PAID ISSUE FEE TOWARD THE ISSUE FEE NOW DUE.

HOW TO REPLY TO THIS NOTICE:

I. Review the ENTITY STATUS shown above. If the ENTITY STATUS is shown as SMALL or MICRO, verify whether entitlement to that entity status still applies.

If the ENTITY STATUS is the same as shown above, pay the TOTAL FEE(S) DUE shown above.

If the ENTITY STATUS is changed from that shown above, on PART B - FEE(S) TRANSMITTAL, complete section number 5 titled "Change in Entity Status (from status indicated above)".

For purposes of this notice, small entity fees are 1/2 the amount of undiscounted fees, and micro entity fees are 1/2 the amount of small entity fees.

II. PART B - FEE(S) TRANSMITTAL, or its equivalent, must be completed and returned to the United States Patent and Trademark Office (USPTO) with your ISSUE FEE and PUBLICATION FEE (if required). If you are charging the fee(s) to your deposit account, section "4b" of Part B - Fee(s) Transmittal should be completed and an extra copy of the form should be submitted. If an equivalent of Part B is filed, a request to reapply a previously paid issue fee must be clearly made, and delays in processing may occur due to the difficulty in recognizing the paper as an equivalent of Part B.

III. All communications regarding this application must give the application number. Please direct all communications prior to issuance to Mail Stop ISSUE FEE unless advised to the contrary.

IMPORTANT REMINDER: Maintenance fees are due in utility patents issuing on applications filed on or after Dec. 12, 1980. It is patentee's responsibility to ensure timely payment of maintenance fees when due. More information is available at www.uspto.gov/PatentMaintenanceFees.

PART B - FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), by mail or fax, or via EFS-Web.

By mail, send to: Mail Stop ISSUE FEE
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

By fax, send to: (571)-273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

86110 7590 05/26/2021
Wolf, Greenfield & Sacks, P.C. (Harvard)
600 Atlantic Avenue
Boston, MA 02210-2206

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being transmitted to the USPTO via EFS-Web or by facsimile to (571) 273-2885, on the date below.

(Typed or printed name)
(Signature)
(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/329,683	01/27/2017	Xiaowei Zhuang	H0498.70509US02	7428

TITLE OF INVENTION: SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	SMALL	$600	$0.00	$0.00	$600	08/26/2021

EXAMINER	ART UNIT	CLASS-SUBCLASS
LU, FRANK WEI MIN	1634	435-091200

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).

❑ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.

❑ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-09 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list
(1) The names of up to 3 registered patent attorneys or agents OR, alternatively,
(2) The name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

1_____
2_____
3_____

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document must have been previously recorded, or filed for recordation, as set forth in 37 CFR 3.11 and 37 CFR 3.81(a). Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE (B) RESIDENCE: (CITY and STATE OR COUNTRY)

Please check the appropriate assignee category or categories (will not be printed on the patent) : ❑ Individual ❑ Corporation or other private group entity ❑ Government

4a. Fees submitted: ❑Issue Fee ❑Publication Fee (if required) ❑Advance Order - # of Copies _____

4b. Method of Payment: *(Please first reapply any previously paid fee shown above)*

❑ Electronic Payment via EFS-Web ❑ Enclosed check ❑ Non-electronic payment by credit card (Attach form PTO-2038)

❑ The Director is hereby authorized to charge the required fee(s), any deficiency, or credit any overpayment to Deposit Account No. _____

5. **Change in Entity Status** (from status indicated above)

❑ Applicant certifying micro entity status. See 37 CFR 1.29

❑ Applicant asserting small entity status. See 37 CFR 1.27

❑ Applicant changing to regular undiscounted fee status.

NOTE: Absent a valid certification of Micro Entity Status (see forms PTO/SB/15A and 15B), issue fee payment in the micro entity amount will not be accepted at the risk of application abandonment.
NOTE: If the application was previously under micro entity status, checking this box will be taken to be a notification of loss of entitlement to micro entity status.
NOTE: Checking this box will be taken to be a notification of loss of entitlement to small or micro entity status, as applicable.

NOTE: This form must be signed in accordance with 37 CFR 1.31 and 1.33. See 37 CFR 1.4 for signature requirements and certifications.

Authorized Signature _____ Date _____

Typed or printed name _____ Registration No. _____

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/329,683	01/27/2017	Xiaowei Zhuang	H0498.70509US02	7428

86110 7590 05/26/2021

Wolf, Greenfield & Sacks, P.C. (Harvard)
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER	
LU, FRANK WEI MIN	
ART UNIT	PAPER NUMBER
1634	

DATE MAILED: 05/26/2021

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)
(Applications filed on or after May 29, 2000)

The Office has discontinued providing a Patent Term Adjustment (PTA) calculation with the Notice of Allowance.

Section 1(h)(2) of the AIA Technical Corrections Act amended 35 U.S.C. 154(b)(3)(B)(i) to eliminate the requirement that the Office provide a patent term adjustment determination with the notice of allowance. See Revisions to Patent Term Adjustment, 78 Fed. Reg. 19416, 19417 (Apr. 1, 2013). Therefore, the Office is no longer providing an initial patent term adjustment determination with the notice of allowance. The Office will continue to provide a patent term adjustment determination with the Issue Notification Letter that is mailed to applicant approximately three weeks prior to the issue date of the patent, and will include the patent term adjustment on the patent. Any request for reconsideration of the patent term adjustment determination (or reinstatement of patent term adjustment) should follow the process outlined in 37 CFR 1.705.

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Customer Service Center of the Office of Patent Publication at 1-(888)-786-0101 or (571)-272-4200.

OMB Clearance and PRA Burden Statement for PTOL-85 Part B

The Paperwork Reduction Act (PRA) of 1995 requires Federal agencies to obtain Office of Management and Budget approval before requesting most types of information from the public. When OMB approves an agency request to collect information from the public, OMB (i) provides a valid OMB Control Number and expiration date for the agency to display on the instrument that will be used to collect the information and (ii) requires the agency to inform the public about the OMB Control Number's legal significance in accordance with 5 CFR 1320.5(b).

The information collected by PTOL-85 Part B is required by 37 CFR 1.311. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 30 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, Virginia 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450. Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Notice of Allowability	Application No. 15/329,683	Applicant(s) Zhuang et al.	
	Examiner FRANK W LU	Art Unit 1634	AIA (FITF) Status Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☑ This communication is responsive to the amendments filed on 1/22/2021 and 3/26/2021.
   ☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2. ☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

3. ☑ The allowed claim(s) is/are 1,71,75,77-78,179-180 and 183-187 . As a result of the allowed claim(s), you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see **http://www.uspto.gov/patents/init_events/pph/index.jsp** or send an inquiry to **PPHfeedback@uspto.gov.**

4. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

   Certified copies:

   a) ☐ All b) ☐ Some *c) ☐ None of the:

   1. ☐ Certified copies of the priority documents have been received.
   2. ☐ Certified copies of the priority documents have been received in Application No. _____ .
   3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

   * Certified copies not received: _____ .

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.
THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.

5. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
   ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date _____ .

   Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).

6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. ☐ Notice of References Cited (PTO-892)
2. ☑ Information Disclosure Statements (PTO/SB/08), Paper No./Mail Date See Continuation Sheet.
3. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material _____.
4. ☑ Interview Summary (PTO-413), Paper No./Mail Date. 4/7/2021.
5. ☑ Examiner's Amendment/Comment
6. ☑ Examiner's Statement of Reasons for Allowance
7. ☐ Other _____.

/FRANK W LU/ Primary Examiner, Art Unit 1634	May 10, 2021

Continuation of Attachment(s) 2. Information Disclosure Statements (PTO/SB/08), Paper No./Mail Date: 4/12/2017, 11/20/2018, 10/31/2019, 12/11/2019, and 1/22/2021

DETAILED ACTION

Reasons for Allowance

1. An examiner's amendment to the record appears below. Should the changes and/or additions be unacceptable to applicant, an amendment may be filed as provided by 37 CFR 1.312. To ensure consideration of such an amendment, it MUST be submitted no later than the payment of the issue fee.

Authorization for this examiner's amendment was given in a telephone interview with Dr. Tani Chen (Reg. No. 52,728) on May 7, 2021.

2. The application has been amended as follows:

Cancel claims 34, 177, 178, and 181.

1. (Currently amended) A method for detecting a plurality of different nucleic acid targets within a sample by *in situ* hybridization, comprising:

(a) producing a plurality of first hybridized complexes within a sample containing a plurality of different nucleic acid targets by exposing [a] the sample [suspected of containing at least some of a plurality of nucleic acid targets] to a plurality of different encoding nucleic acid probes,

wherein [at least some] each of the plurality of encoding nucleic acid probes comprises (i) a target sequence that specifically hybridizes to a different nucleic acid target or a different location of an identical nucleic acid target of [in] the plurality of different nucleic acid targets and (ii) one or more read sequences such that each of the plurality of first hybridized complexes comprises multiple encoding nucleic acid probes of the different encoding nucleic acid probes located in its different locations[,

wherein unique combinations of one or more read sequences correspond to valid

codewords, each of the valid codewords associated to each of the plurality of nucleic acid targets, wherein the valid codeword consists of a series of positions and values for each position, and each read sequence is associated with each of the values, or a subset thereof, at each of the positions of the valid codewords];

(b) [exposing the sample to a first round of one or more readout nucleic acid probes, wherein at least some of the readout nucleic acid probes comprise] producing a plurality of readout probe-hybridized complexes by exposing each of the plurality of first hybridized complexes within the sample to a first readout nucleic acid probe, wherein the first readout nucleic acid probe comprises a first portion and a second portion, the first portion comprising a sequence that specifically hybridizes to one of the read sequences, and the second portion comprising a fluorescent signaling entity capable of producing a fluorescent signal;

(c) [identifying nucleic acid targets in the sample, comprising:]

detecting a fluorescent [signals] signal from each of the plurality of readout probe-hybridized complexes [emitted from the fluorescent signaling entities of bound readout nucleic acid probes in at least some locations within the sample to determine a codeword in at least some locations within the sample based on the read sequences present on the bound encoding nucleic acid probes];

(d) removing the fluorescent signal from each of the plurality of readout probe-hybridized complexes by inactivating the fluorescent signaling entity or removing the fluorescent signaling entity from each of the plurality of readout probe-hybridized complexes;

(e) after step (d), producing a different plurality of readout probe-hybridized complexes by exposing the sample to a different readout nucleic acid probe, wherein the different readout nucleic acid probe comprise a first portion and a second portion, the first portion comprising a

sequence that specifically hybridizes to one of the read sequences, and the second portion comprising a fluorescent signaling entity capable of producing a fluorescent signal;

(f) detecting a fluorescent signal from each of the different plurality of readout probe hybridized complexes;

(g) removing the fluorescent signal from each of the different plurality of readout probe hybridized complexes by inactivating the fluorescent signaling entity or removing the fluorescent signaling entity from each of the different plurality of readout probe hybridized complexes;

(h) repeating steps (e) to (g) for one or more times using a different readout nucleic acid probe, wherein the different readout nucleic acid probe comprises a first portion and a second portion, the first portion comprising a sequence that specifically hybridizes to one of the read sequences, and the second portion comprising a fluorescent signaling entity capable of producing a fluorescent signal;

(i) producing codewords representing the plurality of different nucleic acid targets at locations within the sample, wherein each of the codeword represents one of the plurality of different nucleic acid targets and comprises multiple binary values 1 and 0, wherein a value of 1 is obtained when the fluorescent signal is detected from one of the plurality of readout probe hybridized complexes or one of the different plurality of readout probe-hybridized complexes at a respective location within the sample while a value of 0 is obtained when the fluorescent signal is not detected from one of the plurality of readout probe-hybridized complexes or one of the different plurality of readout probe-hybridized complexes at the respective location within the sample;

(j) matching the codewords with valid codewords in a codebook [identifying the codeword at each location] by comparing the [codeword] codewords to the valid codewords in the

codebook [associated with the plurality of nucleic acid targets], and if [no match is found] one of the codewords is not matched with one of the valid codewords in the codebook, applying an error detection or correction system, matching the one of the codewords with another of the valid codewords in the codebook, [to the codeword to match one of the valid codewords] or discarding [the codeword] the one of the codewords, wherein the codebook comprises the valid codewords of the plurality of nucleic acid targets; and

(k) determining the plurality of nucleic acid targets in the sample based on results from the matching step [assigning the identity of some of the plurality of nucleic acid targets in the sample based on the identity of the nucleic acid targets associated with the valid codewords to which the identified codewords were matched].

71. (Currently Amended) The method of claim 1, [comprising determining the fluorescent signaling entity] wherein the fluorescent signal in step (c) and/or step (f) is detected at a resolution better than 300 nm.

77. (Currently Amended) The method of claim 75, wherein the cell is fixed on a coverslip.

179. (Currently Amended) The method of claim 1, further comprising determining an abundance and/or spatial distribution of the different nucleic acid targets at locations within the sample [based on the codewords matched to the valid codewords in each location within the sample].

180. (Currently Amended) The method of claim 1, further comprising removing [unbound] the readout nucleic acid probes unhybridized to each of the plurality of first hybridized complex prior to step (c) [said detecting fluorescent signals].

184. (Currently Amended) The method of claim 1, wherein in said comparing the [codeword] codewords to the valid codewords in the codebook in step [(c)] (j), one of the codewords is not

matched with one of the valid codewords in the codebook when the [codeword] one of the codewords [differs from the valid codeword compared by] has one or more errors.

185. (Currently Amended) The method of claim 1, wherein in said comparing the [codeword] codewords to the valid codewords in the codebook in step [(c)] (j), one of the codewords is not matched with one of the valid codewords in the codebook when the [codeword] one of the codewords [differs from the valid codeword compared by] has one error.

186. (Currently Amended) The method of claim 1, wherein in said comparing the [codeword] codewords to the valid codewords in the codebook in step [(c)] (j), one of the codewords is not matched with one of the valid codewords in the codebook when the [codeword] one of the codewords [differs from the valid codeword compared by] has two errors.

187. (Currently Amended) The method of claim 1, wherein said comparing the [codeword] codewords to the valid codewords in the codebook in step [(c)] (j) comprises comparing the values at each of the positions of the codewords.

3. The following is an examiner's statement of reasons for allowance:

Claims 1, 71, 75, 77, 78, 179, 180, and 183-187 are allowable in light of the examiner's amendment. The rejections under 35 U.S.C 101, 112 (a) and 112 (b) have been withdrawn in view of the examiner's amendment. No prior art in the record teaches a combination of steps (a) to (k) of claim 1. No prior art either alone or in combination with the other arts in the record teaches/teach or reasonably suggests/suggest a method for detecting a plurality of different nucleic acid targets within a sample by *in situ* hybridization which comprises all limitations recited in claim 1.

Any comments considered necessary by applicant must be submitted no later than

the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance".

4. Papers related to this application may be submitted to Group 1600 by facsimile transmission. Papers should be faxed to Group 1600 via the PTO Fax Center. The faxing of such papers must conform with the notices published in the Official Gazette, 1096 OG 30 (November 15, 1988), 1156 OG 61 (November 16, 1993), and 1157 OG 94 (December 28, 1993)(See 37 CAR § 1.6(d)). The CM Fax Center number is (571)273-8300.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Frank Lu, Ph.D., whose telephone number is (571)272-0746. The examiner can normally be reached on Monday-Friday from 9 A.M. to 5 P.M.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Dave Nguyen, can be reached on (571)272-0731.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/FRANK W LU/
Primary Examiner, Art Unit 1634
May 18, 2021

Examiner-Initiated Interview Summary	Application No. 15/329,683	Applicant(s) Zhuang et al.		
	Examiner FRANK W LU	Art Unit 1634	AIA (First Inventor to File) Status Yes	Page 1 of 1

All Participants (applicant, applicants representative, PTO personnel)	**Title**	**Type**
FRANK W LU	Primary Examiner	Telephonic
Dr. Tani Chen (Reg. No. 52,728)	Attorney of Record	

Date of Interview: 07 May 2021

Issues Discussed:

Proposed Amendment(s)

Dr. Chen agreed with the examiner's amendment.

/FRANK W LU/ Primary Examiner, Art Unit 1634	May 10, 2021

Applicant is reminded that a complete written statement as to the substance of the interview must be made of record in the application file. It is the applicants responsibility to provide the written statement, unless the interview was initiated by the Examiner and the Examiner has indicated that a written summary will be provided. See MPEP 713.04
Please further see:
MPEP 713.04
Title 37 Code of Federal Regulations (CFR) § 1.133 Interviews, paragraph (b)
37 CFR § 1.2 Business to be transacted in writing

Applicant recordation instructions: It is not necessary for applicant to provide a separate record of the substance of interview.

Examiner recordation instructions: Examiners must summarize the substance of any interview of record. A complete and proper recordation of the substance of an interview should include the items listed in MPEP 713.04 for complete and proper recordation including the identification of the general thrust of each argument or issue discussed, a general indication of any other pertinent matters discussed regarding patentability and the general results or outcome of the interview, to include an indication as to whether or not agreement was reached on the issues raised.

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15/329,683	Zhuang et al.
	Examiner	Art Unit
	FRANK W LU	1634

CPC						
Symbol					Type	Version
C12N	/	15	/	1065	F	2013-01-01
C12N	/	15	/	10	I	2013-01-01
C07H	/	21	/	02	I	2013-01-01
C07H	/	21	/	04	I	2013-01-01
G16B	/	25	/	00	I	2019-02-01
C12Q	/	1	/	6806	I	2013-01-01
C12N	/	15	/	1093	I	2013-01-01
C12Q	/	1	/	6837	I	2013-01-01
G16B	/	30	/	00	A	2019-02-01
G16B	/	40	/	00	A	2019-02-01
C12Q	/	1	/	6816	A	2013-01-01
C12Q	/	1	/	6841	A	2013-01-01
C12Q	/	1	/	6869	A	2013-01-01
G06N	/	7	/	005	A	2013-01-01

CPC Combination Sets								
Symbol					Type	Set	Ranking	Version
C12Q	/	1	/	6806	I	1	1	2013-01-01
C12Q	/	2521	/	107	I	1	2	2013-01-01
C12Q	/	2525	/	143	I	1	3	2013-01-01
C12Q	/	2525	/	179	I	1	4	2013-01-01
C12Q	/	2537	/	143	I	1	5	2013-01-01
C12Q	/	2565	/	514	I	1	6	2013-01-01
C12Q	/	1	/	6816	I	2	1	2013-01-01
C12Q	/	2525	/	161	I	2	2	2013-01-01
C12Q	/	2537	/	143	I	2	3	2013-01-01
C12Q	/	2563	/	179	I	2	4	2013-01-01
C12Q	/	1	/	6816	I	3	1	2013-01-01
C12Q	/	2525	/	161	I	3	2	2013-01-01
C12Q	/	2537	/	143	I	3	3	2013-01-01

NONE (Assistant Examiner)	(Date)	Total Claims Allowed: 12	
/FRANK W LU/ Primary Examiner, Art Unit 1634 (Primary Examiner)	10 May 2021 (Date)	O.G. Print Claim(s) 1	O.G. Print Figure 3A

U.S. Patent and Trademark Office — Part of Paper No.: 5/2021

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15/329,683	Zhuang et al.
	Examiner	**Art Unit**
	FRANK W LU	1634

CPC Combination Sets								
Symbol					**Type**	**Set**	**Ranking**	**Version**
C12Q	/	2563	/	179	I	3	4	2013-01-01
C12Q	/	2565	/	102	I	3	5	2013-01-01

NONE		**Total Claims Allowed:**	
(Assistant Examiner)	(Date)	12	
/FRANK W LU/ Primary Examiner, Art Unit 1634	10 May 2021	O.G. Print Claim(s)	O.G. Print Figure
(Primary Examiner)	(Date)	1	3A

U.S. Patent and Trademark Office — Part of Paper No.: 5/2021

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15/329,683	Zhuang et al.
	Examiner	**Art Unit**
	FRANK W LU	1634

INTERNATIONAL CLASSIFICATION				
CLAIMED				
C12Q	/	1	/	68

NON-CLAIMED				
	/		/	

US ORIGINAL CLASSIFICATION	
CLASS	**SUBCLASS**
435	91.2

CROSS REFERENCES(S)						
CLASS	**SUBCLASS (ONE SUBCLASS PER BLOCK)**					

NONE (Assistant Examiner)	(Date)	**Total Claims Allowed:** 12	
/FRANK W LU/ Primary Examiner, Art Unit 1634 (Primary Examiner)	10 May 2021 (Date)	O.G. Print Claim(s) 1	O.G. Print Figure 3A

U.S. Patent and Trademark Office — Part of Paper No.: 5/2021

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15/329,683	Zhuang et al.
	Examiner	**Art Unit**
	FRANK W LU	1634

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47

CLAIMS

Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original
1	1	10	185												
2	71	11	186												
3	75	12	187												
4	77														
5	78														
6	179														
7	180														
8	183														
9	184														

NONE		**Total Claims Allowed:**	
(Assistant Examiner)	(Date)	12	
/FRANK W LU/ Primary Examiner, Art Unit 1634	10 May 2021	O.G. Print Claim(s)	O.G. Print Figure
(Primary Examiner)	(Date)	1	3A

U.S. Patent and Trademark Office — Part of Paper No.: 5/2021

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15/329,683	Zhuang et al.
	Examiner	Art Unit
	FRANK W LU	1634

CPC - Searched*		
Symbol	**Date**	**Examiner**
Update	05/07/2021	/FL/

CPC Combination Sets - Searched*		
Symbol	**Date**	**Examiner**
Update	05/07/2021	/FL/

US Classification - Searched*			
Class	**Subclass**	**Date**	**Examiner**
Update		05/07/2021	/FL/

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

Search Notes		
Search Notes	**Date**	**Examiner**
Update (see attached files)	05/07/2021	/FL/

/FRANK W LU/ Primary Examiner, Art Unit 1634	May 10, 2021

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15/329,683	Zhuang et al.
	Examiner	Art Unit
	FRANK W LU	1634

Interference Search			
US Class/CPC Symbol	**US Subclass/CPC Group**	**Date**	**Examiner**
CPC C12Q	1/6806, 6816, 6837, 6841, 6869	05/07/2021	/FL/
CPC C12N	15/10, 1065, 1093	05/07/2021	/FL/
CPC G16B	25/00; 30/00; 40/00	05/07/2021	/FL/
CPC C07H	21/02, 04	05/07/2021	/FL/
CPC G06N	7/05	05/07/2021	/FL/

/FRANK W LU/ Primary Examiner, Art Unit 1634	May 10, 2021

PART II: Compliance with 37 C.F.R. § 1.704(d)

Each item of information contained in this Information Disclosure Statement

(i) was first cited in any communication from a patent office in a counterpart foreign or international application or from the United States Patent and Trademark Office, and this communication was not received by any individual designated in 37 C.F.R. § 1.56(c) more than thirty (30) days prior to the filing of the Information Disclosure Statement; or

(ii) is a communication that was issued by a patent office in a counterpart foreign or international application or by the United States Patent and Trademark Office, and this communication was not received by any individual designated in 37 C.F.R. § 1.56(c) more than thirty (30) days prior to the filing of the Information Disclosure Statement.

PART III: Information Cited

The undersigned hereby makes of record in the above-identified application the information listed on the attached form PTO-1449 (modified PTO/SB/08). The order of presentation of the references should not be construed as an indication of the importance of the references.

The undersigned hereby makes the following additional information of record in the above-identified application.

The undersigned would like to bring to the Examiner's attention the enclosed search report or other communication from a corresponding or related International or Foreign National Application:

Serial No.	Date of Mailing	Type(s) of Communication	Docket No.
CN 201580052678.7	**11-14-2019**	**Chinese Office Action**	**H0498.70509CN00**

PART IV: Remarks

Documents cited anywhere in the Information Disclosure Statement are enclosed unless otherwise indicated. It is respectfully requested that:

1. The Examiner consider completely the cited information, along with any other information, in reaching a determination concerning the patentability of the present claims;

/FRANK W LU/ 05/07/2021

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

FORM PTO-1449/A and B (modified PTO/SB/08)				APPLICATION NO.: 15/329,683	ATTY. DOCKET NO.: H0498.70509US02
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				FILING DATE: January 27, 2017	CONFIRMATION NO.: 7428
				FIRST NAMED INVENTOR: Xiaowei Zhuang	
Sheet	1	of	2	GROUP ART UNIT: 1634	EXAMINER: F. W. Lu

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document		Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		Number	Kind Code		
		2015-144490	A1	Deisseroth et al.	05-28-2015
		2016-0153005	A1	Cong et al.	06-02-2016
		2018-0230226	A1	Podack et al.	08-16-2018

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document			Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)
		Office/ Country	Number	Kind Code			
		WO	2018/089438	A1	President and Fellows of Harvard College	05-17-2018	

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		European Office Action mailed October 16, 2020 for Application No. 15827358.1 (H0498.70509EP00).	
		Chinese Office Action dated August 19, 2020 for Application No. CN 201580051029.5 (H0498.70510CN00).	Y - Full
		European Office Action mailed August 19, 2020 for Application No. 15828133.7 (H0498.70510EP00).	
		Extended European Search Report mailed May 20, 2020 for Application No. 17869122.6 (H0498.70630EP00).	
		Supplementary European Search Report mailed June 9, 2020 for Application No. 17869122.6 (H0498.70630EP00).	
		Extended European Search Report mailed July 20, 2020 for Application No. 17869515.1 (H0498.70645EP00).	
		European Office Action mailed August 6, 2020 for Application No. 17869515.1 (H0498.70645EP00).	
		International Search Report and Written Opinion for Application No. PCT/US2019/065857 mailed May 1, 2020 (H0498.70680WO00).	
		International Search Report and Written Opinion for Application No. PCT/US20/28632 mailed July 16, 2020 (H0498.70693WO00).	
		CHEN et al., RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science. 2015;348(6233):aaa6090.	
		CHEN et al., Supplemental Material for Expansion microscopy. Science. 2015 Jan; 347(6221):1-18.	
		CHUNG et al., CLARITY for mapping the nervous system [published correction appears in Nat Methods. 2013 Oct;10(10):1035]. Nat Methods. 2013;10(6):508-513.	
		TSANOV et al., smiFISH and FISH-quant - a flexible single RNA detection approach with super-resolution capability. Nucleic Acids Res. 2016;44(22):e165.	
		YANG et al., Single-cell phenotyping within transparent intact tissue through whole-body clearing. Cell. 2014;158(4):945-958.	

EXAMINER:	DATE CONSIDERED:
/FRANK W LU/	05/07/2021

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

8892647.1

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

FORM PTO-1449/A and B (modified PTO/SB/08)				APPLICATION NO.: 15/329,683	ATTY. DOCKET NO.: H0498.70509US02
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				FILING DATE: January 27, 2017	CONFIRMATION NO.: 7428
				FIRST NAMED INVENTOR: Xiaowei Zhuang	
Sheet	2	of	2	GROUP ART UNIT: 1634	EXAMINER: F. W. Lu

		ZHANG et al., Tandem Spinach Array for mRNA Imaging in Living Bacterial Cells. Sci Rep. 2015;5:17295.	

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

EXAMINER:	DATE CONSIDERED:
/FRANK W LU/	05/18/2021

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

8892647.1

PART II: Information Cited

The undersigned hereby makes of record in the above-identified application the information listed on the attached form PTO-1449 (modified PTO/SB/08). The order of presentation of the references should not be construed as an indication of the importance of the references.

The undersigned hereby makes the following additional information of record in the above-identified application.

The undersigned would like to bring to the Examiner's attention the following co-pending applications that may contain subject matter related to this application:

Serial No.	Filing Date	Inventor(s)	Docket No.
*16/253,919	1/22/2019	Zhuang et al.	H0498.70510US02
*16/347,874	5/7/2019	Zhuang et al.	H0498.70630US01
*16/616,833	11/25/2019	Zhuang et al.	H0498.70637US01
*16/348,071	5/7/2019	Zhuang et al.	H0498.70645US01

*A copy of this reference is not provided as the Office has waived the requirement under 37 C.F.R. § 1.98(a)(2)(iii) for submitting a copy of a cited U.S. patent application if it is scanned to the Image File Wrapper system and is available on Private PAIR .

The undersigned would like to bring to the Examiner's attention the enclosed search report or other communication from a corresponding or related International or Foreign National Application:

Serial No.	Date of Mailing	Type(s) of Communication	Docket No.
EP 15827358.1	**10-16-2020**	**European Office Action**	**H0498.70509EP00**
CN 201580051029.5	**08-19-2020**	**Chinese Office Action**	**H0498.70510CN00**
EP 15828133.7	**08-19-2020**	**European Office Action**	**H0498.70510EP00**
EP 17869122.6	**05-20-2020**	**Extended European Search Report**	**H0498.70630EP00**
EP 17869122.6	**06-09-2020**	**Supplementary European Search Report**	**H0498.70630EP00**
EP 17869515.1	**07-20-2020**	**Extended European Search Report**	**H0498.70645EP00**
EP 17869515.1	**08-06-2020**	**European Office Action**	**H0498.70645EP00**
PCT/US2019/065857	**05-01-2020**	**International Search Report and Written Opinion**	**H0498.70680WO00**
PCT/US20/28632	**07-16-2020**	**International Search Report and Written Opinion**	**H0498.70693WO00**

/FRANK W LU/ 05/07/2021

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

presentation of the references should not be construed as an indication of the importance of the references.

The undersigned would like to bring to the Examiner's attention the following co-pending applications that may contain subject matter related to this application:

Serial No.	Filing Date	Inventor(s)	Docket No.
* 15/329,651	01-27-2017	Zhuang et al.	H0498.70510US01

*A copy of this reference is not provided as the Office has waived the requirement under 37 C.F.R. § 1.98(a)(2)(iii) for submitting a copy of a cited U.S. patent application if it is scanned to the Image File Wrapper system and is available on Private PAIR.

The undersigned would like to bring to the Examiner's attention the enclosed search report or other communication from a corresponding or related International or Foreign National Application:

Serial No.	Date of Mailing	Type(s) of Communication	Docket No.
PCT/US2015/042556	**10-06-2015**	**Invitation to Pay Additional Fees**	**H0498.70509WO00**
PCT/US2015/042556	**12-18-2015**	**International Search Report and Written Opinion**	**H0498.70509WO00**
PCT/US2015/042556	**02-09-2017**	**International Preliminary Report on Patentability**	**H0498.70509WO00**
PCT/US2015/042559	**10-09-2015**	**International Search Report and Written Opinion**	**H0498.70510WO00**
PCT/US2015/042559	**02-09-2017**	**International Preliminary Report on Patentability**	**H0498.70510WO00**

PART III: Remarks

Documents cited anywhere in the Information Disclosure Statement are enclosed unless otherwise indicated. It is respectfully requested that:

1. The Examiner consider completely the cited information, along with any other information, in reaching a determination concerning the patentability of the present claims;

2. The enclosed form PTO-1449 (modified PTO/SB/08) be signed by the Examiner to evidence that the cited information has been fully considered by the United States Patent and Trademark Office during the examination of this application;

/FRANK W LU/ 05/07/2021

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	0	codeword same binary near value same codebook same (hybridize or hybridizing or hybridization) same ("nucleic acid" or DNA or RNA or polynucleotide) same (fluorescence or fluorophore or fluorescent)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/05/07 20:56
L2	0	(barcode or codeword) same binary near value same codebook same (hybridize or hybridizing or hybridization) same ("nucleic acid" or DNA or RNA or polynucleotide) same (fluorescence or fluorophore or fluorescent)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/05/07 20:57
L3	1	(barcode or codeword) same codebook same (hybridize or hybridizing or hybridization) same ("nucleic acid" or DNA or RNA or polynucleotide) same (fluorescence or fluorophore or fluorescent)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/05/07 20:57
L4	0	(barcode or codeword) same binary near value same (hybridize or hybridizing or hybridization) same ("nucleic acid" or DNA or RNA or polynucleotide) same (fluorescence or fluorophore or fluorescent)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/05/07 20:58
L5	833	(barcode or codeword) same (hybridize or hybridizing or hybridization) same ("nucleic acid" or DNA or RNA or polynucleotide) same (fluorescence or fluorophore or fluorescent)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/05/07 20:58
L6	20	(barcode or codeword) same (hybridize or hybridizing or hybridization) same ("nucleic acid" or DNA or RNA or polynucleotide) same (fluorescence or fluorophore or fluorescent) same (correct or correcting or correction)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/05/07 20:59
L7	2	(barcode or codeword) same (hybridize or hybridizing or hybridization) same ("nucleic acid" or DNA or RNA or polynucleotide) same (fluorescence or fluorophore or fluorescent) same (correct or correcting or correction) same (match or matching)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/05/07 20:59

L8	16	6 AND ((C12Q2563/179 OR C12Q2565/518 OR C12Q1/6841 OR C12Q2563/107 OR C12Q2565/514 OR C12Q1/6869 OR C12Q2565/101 OR C12Q2537/149 OR C12Q1/6874 OR C12Q2525/161 OR C12Q1/6816 OR C12Q2537/143 OR C12Q2537/155 OR C12Q1/6837 OR C12Q1/6827 OR C12Q2533/107 OR C12Q2565/102 OR C12Q1/6806 OR C12Q2521/301 OR C12Q2521/319 OR C12Q1/682 OR C12Q1/6825 OR C12Q2521/501 OR C12Q2565/631 OR C12Q1/6804 OR C12Q1/6876 OR C12Q1/6881 OR C12Q2521/107 OR C12Q2522/101 OR C12Q2525/113 OR C12Q2525/143 OR C12Q2525/179 OR C12Q2525/301 OR C12Q2535/122 OR C12Q2545/101 OR C12Q2600/158 OR C12Q1/68 OR C12Q1/6886 OR C12Q2535/131 OR C12Q2537/125 OR C12Q2600/118 OR G16B25/00 OR G16B40/00 OR G16B30/00 OR G01N21/6458 OR G01N21/78 OR G01N1/36 OR G01N2440/00 OR G01N33/48721 OR G01N33/6875 OR C07H21/02 OR C07H21/04 OR C12N15/10 OR C12N15/1034 OR C12N15/1058 OR C12N15/1065 OR C12N15/1079 OR C12N15/1086 OR C12N15/1093 OR C12N15/115 OR C12N2310/16 OR C12N2320/10 OR C40B30/04 OR C40B70/00 OR G06N7/005).CPC.)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/05/07 21:01

5/7/2021 9:02:15 PM
C:\Users\flu\Documents\EAST\Workspaces\~auto.01112018.114853.bak.wsp

PART II: Compliance with 37 C.F.R. § 1.704(d)

Each item of information contained in this Information Disclosure Statement

(i) was first cited in any communication from a patent office in a counterpart foreign or international application or from the United States Patent and Trademark Office, and this communication was not received by any individual designated in 37 C.F.R. § 1.56(c) more than thirty (30) days prior to the filing of the Information Disclosure Statement; or

(ii) is a communication that was issued by a patent office in a counterpart foreign or international application or by the United States Patent and Trademark Office, and this communication was not received by any individual designated in 37 C.F.R. § 1.56(c) more than thirty (30) days prior to the filing of the Information Disclosure Statement.

PART III: Information Cited

The undersigned hereby makes of record in the above-identified application the information listed on the attached form PTO-1449 (modified PTO/SB/08). The order of presentation of the references should not be construed as an indication of the importance of the references.

The undersigned hereby makes the following additional information of record in the above-identified application.

The undersigned would like to bring to the Examiner's attention the enclosed search report or other communication from a corresponding or related International or Foreign National Application:

Serial No.	Date of Mailing	Type(s) of Communication	Docket No.
EP 15827358.1	**10-16-2019**	**European Office Action**	**H0498.70509EP00**

PART IV: Remarks

Documents cited anywhere in the Information Disclosure Statement are enclosed unless otherwise indicated. It is respectfully requested that:

/FRANK W LU/ 05/07/2021

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

```
 (FILE 'HOME' ENTERED AT 21:03:02 ON 07 MAY 2021)

     FILE 'MEDLINE, CAPLUS, BIOSIS, SCISEARCH, EMBASE' ENTERED AT 21:03:26 ON
     07 MAY 2021
L1         45100 S BARCODE OR CODEWORD
L2         45100 S BARCODE OR CODEWORD
L3       1436575 S HYBRIDIZE OR HYBRIDIZING OR HYBRIDIZATION
L4      13702939 S  NUCLEIC ACID OR DNA OR RNA OR POLYNUCLEOTIDE
L5       3969546 S FLUORESCENCE OR FLUOROPHORE OR FLUORESCENT
L6       2539155 S CORRECT OR CORRECTING OR CORRECTION
L7       1115639 S MATCH OR MATCHING
L8             0 S L2 AND L3 AND L4 AND L5 AND L6 AND L7
L9            21 S L2 AND L3 AND L4 AND L5 AND L6
L10            7 DUP REM L9 (14 DUPLICATES REMOVED)
```

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/329,683	01/27/2017	Xiaowei Zhuang	H0498.70509US02	7428

86110 7590 06/04/2021
Wolf, Greenfield & Sacks, P.C. (Harvard)
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER
LU, FRANK WEI MIN

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
06/04/2021	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

Patents_eOfficeAction@WolfGreenfield.com
H0498_eOfficeAction@WolfGreenfield.com

UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

Application No. : 15329683
Applicant : Zhuang
Filing Date : 01/27/2017
Date Mailed : 06/04/2021

NOTICE TO FILE CORRECTED APPLICATION PAPERS

Notice of Allowance Mailed

This application has been accorded an Allowance Date and is being prepared for issuance. The application, however, is incomplete for the reasons below.

Applicant is given two (2) months from the mail date of this Notice within which to respond. This time period for reply is extendable under 37 CFR 1.136(a) for only TWO additional MONTHS.

The informalities requiring correction are indicated in the attachment(s). If the informality pertains to the abstract, specification (including claims) or drawings, the informality must be corrected with an amendment in compliance with 37 CFR 1.121 (or, if the application is a reissue application, 37 CFR 1.173). Such an amendment may be filed after payment of the issue fee if limited to correction of informalities noted herein. See Waiver of 37 CFR 1.312 for Documents Required by the Office of Patent Publication, 1280 Off. Gaz. Patent Office 918 (March 23, 2004). In addition, if the informality is not corrected until after payment of the issue fee, for purposes of 35 U.S.C. 154(b)(1)(iv), "all outstanding requirements" will be considered to have been satisfied when the informality has been corrected. A failure to respond within the above-identified time period will result in the application being ABANDONED.

See attachment(s).

A copy of this notice MUST be returned with the reply. Please address response to "Mail Stop Issue Fee, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450".

/Kam Sin/
Publication Branch
Office of Data Management
(571) 272-4200

Application No. 15329683

SPECIFICATION NOT IN COMPLIANCE WITH 37 CFR 1.52(b)(5)

The pages of the specification document coded SPEC dated 07/29/2019 have not been numbered. Per 37 CFR 1.52(b)(5), "the pages of the specification including claims and abstract must be numbered consecutively, beginning with 1, the numbers being centrally located above or preferably below, the text." In response to this notice, the applicant must submit a substitute specification in which the pages are so numbered.

NOTE: Although 37 CFR 1.52(b)(5) refers to page numbering for "the specification including claims and abstract," any abstract or claims submitted in response to this notice will not be entered. Only the substitute specification, and any amendment thereto entered during prosecution, will be entered.

COPY

UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

Application No. : 15329683
Applicant : Zhuang
Filing Date : 01/27/2017
Date Mailed : 06/04/2021

NOTICE TO FILE CORRECTED APPLICATION PAPERS

Notice of Allowance Mailed

This application has been accorded an Allowance Date and is being prepared for issuance. The application, however, is incomplete for the reasons below.

Applicant is given two (2) months from the mail date of this Notice within which to respond. This time period for reply is extendable under 37 CFR 1.136(a) for only TWO additional MONTHS.

The informalities requiring correction are indicated in the attachment(s). If the informality pertains to the abstract, specification (including claims) or drawings, the informality must be corrected with an amendment in compliance with 37 CFR 1.121 (or, if the application is a reissue application, 37 CFR 1.173). Such an amendment may be filed after payment of the issue fee if limited to correction of informalities noted herein. See Waiver of 37 CFR 1.312 for Documents Required by the Office of Patent Publication, 1280 Off. Gaz. Patent Office 918 (March 23, 2004). In addition, if the informality is not corrected until after payment of the issue fee, for purposes of 35 U.S.C. 154(b)(1)(iv), "all outstanding requirements" will be considered to have been satisfied when the informality has been corrected. A failure to respond within the above-identified time period will result in the application being ABANDONED.

See attachment(s).

A copy of this notice MUST be returned with the reply. Please address response to "Mail Stop Issue Fee, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450".

/Kam Sin/
Publication Branch
Office of Data Management
(571) 272-4200

COPY

Application No. 15329683

SPECIFICATION NOT IN COMPLIANCE WITH 37 CFR 1.52(b)(5)

The pages of the specification document coded SPEC dated 07/29/2019 have not been numbered. Per 37 CFR 1.52(b)(5), "the pages of the specification including claims and abstract must be numbered consecutively, beginning with 1, the numbers being centrally located above or preferably below, the text." In response to this notice, the applicant must submit a substitute specification in which the pages are so numbered.

NOTE: Although 37 CFR 1.52(b)(5) refers to page numbering for "the specification including claims and abstract," any abstract or claims submitted in response to this notice will not be entered. Only the substitute specification, and any amendment thereto entered during prosecution, will be entered.

Electronic Patent Application Fee Transmittal

Application Number:	15329683
Filing Date:	27-Jan-2017
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Filer:	Tani Chen/Angela Griffith
Attorney Docket Number:	H0498.70509US02

Filed as Small Entity

Filing Fees for U.S. National Stage under 35 USC 371

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
UTILITY APPL ISSUE FEE	2501	1	600	600

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension-of-Time:				
Miscellaneous:				
			Total in USD ($)	**600**

Electronic Acknowledgement Receipt

EFS ID:	43233443
Application Number:	15329683
International Application Number:	
Confirmation Number:	7428
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Customer Number:	86110
Filer:	Tani Chen/Angela Griffith
Filer Authorized By:	Tani Chen
Attorney Docket Number:	H0498.70509US02
Receipt Date:	13-JUL-2021
Filing Date:	27-JAN-2017
Time Stamp:	12:15:24
Application Type:	U.S. National Stage under 35 USC 371

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$600
RAM confirmation Number	E20217CC15434843
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Transmittal Letter	H049870509US02-TRN-TC.pdf	26287 c7f958517d8ef05e05b14438e0778388e453bcdf	no	1
Warnings:					
Information:					
2	Issue Fee Payment (PTO-85B)	H049870509US02-ISSFEE-TC.pdf	38068 0613b1298ce14f22e69e094f64655d814a16bf56	no	1
Warnings:					
Information:					
3	Specification	H049870509US02-SUBSPEC-TC.pdf	334975 61d21874d357a92b856366629b5630d2ea844c3f	no	90
Warnings:					
Information:					
4	Specification	H049870509US02-MUSPEC-TC.pdf	338034 02a45a50746a916df5c6bb7908b29d19effeca1f	no	90
Warnings:					
Information:					
5	Miscellaneous Incoming Letter	H049870509US02-CCAP-TC.PDF	82773 a70197cb47336d5adb0100ecc14829b4d9c58c56	no	2
Warnings:					
Information:					
6	Fee Worksheet (SB06)	fee-info.pdf	30737 16de560df46a8c6136ceb7fd0ebf26dc6fe1bc17	no	2
Warnings:					
Information:					

Total Files Size (in bytes):	850874

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/329,683-Conf. #7428
Filing Date	January 27, 2017
First Named Inventor	Xiaowei Zhuang
Art Unit	1634
Examiner Name	F. W. Lu
Attorney Docket Number	H0498.70509US02

ENCLOSURES (*Check all that apply*)

- [ ] Fee Transmittal Form
  - [ ] Fee Attached
- [ ] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [ ] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) (please Identify below):
  - Part B – Issue Fee
  - Substitute Specification (clean)
  - Substitute Specification (marked up)
  - Copy of Notice to File Corrected Application Papers

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Tani Chen/		
Printed name	Tani Chen, Sc.D.		
Date	July 13, 2021	Reg. No.	52,728

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: July 13, 2021 Electronic Signature for Angela M. Griffith: /Angela M. Griffith/

9396161.1

PART B–FEE(S) TRANSMITTAL

Complete and send this form, together with the applicable fee(s), by mail or fax, or via EFS-Web.

By mail, send to: Mail Stop ISSUE FEE
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

By fax, send to: (571) 273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being transmitted to the USPTO via EFS-Web or by facsimile to (571) 273-2885, on the date below.

Angela M. Griffith	(Typed or printed name)
/Angela M. Griffith/	(Signature)
July 13, 2021	(Date)

APPLICATION NO	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO	CONFIRMATION NO
15/329,683	01/27/2017	Xiaowei Zhuang	H0498.70509US02	7428

TITLE OF INVENTION: SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

APPLN: TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	SMALL	$600.00			$600.00	08/26/2021

EXAMINER	ART UNIT	CLASS-SUBCLASS
F. W. Lu	1634	

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363)

☐ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.

☐ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-09 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list

(1) The names of up to 3 registered patent attorneys or agents OR, alternatively,

(2) The name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

1. Wolf, Greenfield & Sacks, P.C.
2.
3.

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document must have been previously recorded, or filed for recordation, as set forth in 37 CFR 3.11 and 37 CFR 3.81(a). Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE: President and Fellows of Harvard College

(B) RESIDENCE: (CITY and STATE or COUNTRY): Cambridge, Massachusetts

Please check the appropriate assignee category or categories (will not be printed on the patent): ☐ Individual ☒ Corporation or other private group entity ☐ Government

4a. Fees Submitted: ☒ Issue Fee ☐ Publication Fee (if required) ☐ Advance Order - # of Copies _____

4b. Method of Payment (*Please first reapply any previously paid fee shown above*):

☒ Electronic Payment via EFS-Web ☐ Enclosed check ☐ Non-electronic payment by credit card (Attach form PTO-2038)

☒ The Director is hereby authorized to charge ~~the required fee(s),~~ any deficiency, or credit any overpayment to Deposit Account No. 23/2825

5. Change of Entity Status (from status indicated above)

☐ Applicant certifying micro entity status. See 37 CFR 1.29. NOTE: Absent a valid Certification of Micro Entity Status (see forms PTO/SB/15A and 15B), issue fee payment in the micro entity amount will not be accepted at the risk of application abandonment.

☐ Applicant asserting small entity status. See 37 CFR 1.27. NOTE: If the application was previously under micro entity status, checking this box will be taken as a notification of loss of entitlement to micro entity status.

☐ Applicant changing to regular undiscounted fee status. NOTE: Checking this box will be taken as a notification of loss of entitlement to small or micro entity status, as applicable.

NOTE: This form must be signed in accordance with 37 CFR 1.31 and 1.33. See 37 CFR 1.4 for signature requirements and certifications.

Authorized Signature /Tani Chen/ Date July 13, 2021

Typed or printed name Tani Chen, Sc.D. Registration No. 52,728

9396162.1

SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

RELATED APPLICATIONS

This application is a national stage filing of International Patent Application Serial No. PCT/US2015/042556, filed July 29, 2017, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*, which claims the benefit of U.S. Provisional Patent Application Serial No. 62/031,062, filed July 30, 2014, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/142,653, filed April 3, 2015, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; and U.S. Provisional Patent Application Serial No. 62/050,636, filed September 15, 2014, entitled "Probe Library Construction," by Zhuang, *et al.* Each of the above is incorporated herein by reference.

GOVERNMENT FUNDING

This invention was made with government support under Grant No. GM096450 awarded by National Institutes of Health. The government has certain rights in the invention.

FIELD

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined.

BACKGROUND

Single-molecule fluorescent in situ hybridization (smFISH) is a powerful method for detecting individual mRNA molecules in cells. The high detection efficiency and large dynamic range of this method provides exquisite detail into the expression state, spatial distribution within cells and intact tissues, and variation among cells of individual mRNAs. Such approaches have been essential to many recent insights into understanding gene regulation and expression. A fundamental limitation of smFISH, however, is its low throughput, typically only a few genes at a time. This low throughput is due to a lack of distinguishable probes with which to label cells and the cost of producing large amounts of labeled probe required for high efficiency staining. Thus, improvements in detecting mRNA molecules are needed.

SUMMARY

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. The subject matter of the present invention

involves, in some cases, interrelated products, alternative solutions to a particular problem, and/or a plurality of different uses of one or more systems and/or articles.

In one aspect, the present invention is generally directed to a composition. According to one set of embodiments, the composition comprises a plurality of nucleic acid probes, at least some of which comprise a first portion comprising a target sequence and a plurality of read sequences. In some cases, each comprises a first portion comprising a target sequence and a plurality of read sequences. In some embodiments, the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

In another aspect, the present invention is generally directed to a method. In one set of embodiments, the method includes acts of exposing a sample to a plurality of nucleic acid probes; for each of the nucleic acid probes, determining binding of the nucleic acid probes within the sample; creating codewords based on the binding of the nucleic acid probes; and for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

The method, in another set of embodiments, includes acts of exposing a sample to a plurality of nucleic acid probes, wherein the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences, and wherein at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences; and for each of the nucleic acid probes, determining binding of the target sequences of the nucleic acid probes within the sample.

In yet another set of embodiments, the method includes acts of exposing a sample to a plurality of primary nucleic acid probes (also called encoding probes); exposing the plurality of primary nucleic acid probes to a sequence of secondary nucleic acid probes (also called readout probes) and determining fluorescence of each of the secondary nucleic acid probes within the sample; creating codewords based on fluorescence of the secondary nucleic acid probes; and for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

In one set of embodiments, the method includes acts of exposing a plurality of primary nucleic acid probes to a sample; and exposing the plurality of nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary probes within the sample. In some embodiments, at least some of the plurality of secondary nucleic acid probes comprises distinguishable secondary nucleic acid probes

formed from combinatorial combination of one or more read sequences (or readout probe sequences) taken from a plurality of read sequences (or readout probe sequences).

In another set of embodiments, the method comprises acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a precision better than 500 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a precision better than 500 nm.

In another set of embodiments, the method comprises acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a resolution better than 100 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a resolution better than 100 nm.

The method, in yet another set of embodiments, includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity using a super-resolution imaging technique, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In certain embodiments, the method comprises acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords; and forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

In addition, in one set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes each comprising a target sequence; and forming groups comprising the

plurality of nucleic acid probes such that each group of nucleic acid probes corresponds to at least one common value of a position within the codewords.

In another set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions that is less than the number of targets, and wherein each codeword is associated with a single target, associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords, and forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

The method, in still another set of embodiments, includes acts of exposing a plurality of nucleic acid probes to a cell, exposing the plurality of nucleic acid probes to a sequence of secondary probes and determining fluorescence of each of the secondary probes within the cell, and based on the sequence of fluorescence of each of the secondary probes, determining nucleic acids within the cell.

In another set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes each comprising a target sequence; and forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes correspond to at least one common value of a position within the codewords.

In yet another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity using a super-resolution imaging technique, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In still another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a precision better than 500 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a precision better than 500 nm.

In still another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a resolution better than 100 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In still another set of embodiments, the method includes acts of associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code; associating unique read sequences with each possible value of each position in the codewords, wherein the read sequences are taken from a set of orthogonal sequences, which have limited homology with one another and with the nucleic acid species in a sample; forming a plurality of primary nucleic acid probes each comprising a target sequence that uniquely binds to a nucleic acid target and one or more read sequences; forming a plurality of secondary nucleic acid probes comprising a signaling entity and a sequence that is complementary to one of the read sequences; exposing a sample to the primary nucleic acid probes such that the nucleic acid probes hybridize to the nucleic acid targets in the sample; exposing the primary nucleic acid probes in the sample to a secondary nucleic acid probe such that the secondary nucleic acid probe hybridizes to the read sequence on at least some of the primary nucleic acid probes; imaging the sample; and repeating the exposing and imaging steps one or more times, using a different secondary nucleic acid probe for at least some of the repetitions.

The method, according to yet another set of embodiments, includes acts of associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes comprising a signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets; grouping the nucleic acid probes into a plurality of probe pools, wherein each of the probe pools corresponds to a specific value of a unique position within the codewords; exposing a sample to one of the probe pools; imaging the sample; and repeating the exposing and imaging steps one or more times, using a different probe pool for at least some of the repetitions.

In another aspect, the present invention encompasses methods of making one or more of the embodiments described herein. In still another aspect, the present invention encompasses methods of using one or more of the embodiments described herein.

Other advantages and novel features of the present invention will become apparent from the following detailed description of various non-limiting embodiments of the invention when considered in conjunction with the accompanying figures. In cases where the present specification and a document incorporated by reference include conflicting and/or inconsistent disclosure, the present specification shall control. If two or more documents incorporated by reference include conflicting and/or inconsistent disclosure with respect to each other, then the document having the later effective date shall control.

BRIEF DESCRIPTION OF THE DRAWINGS

Non-limiting embodiments of the present invention will be described by way of example with reference to the accompanying figures, which are schematic and are not intended to be drawn to scale. In the figures, each identical or nearly identical component illustrated is typically represented by a single numeral. For purposes of clarity, not every component is labeled in every figure, nor is every component of each embodiment of the invention shown where illustration is not necessary to allow those of ordinary skill in the art to understand the invention. In the figures:

Figs. 1A-1C illustrates an encoding scheme for nucleic acid probes, in certain embodiments of the invention;

Figs. 2A-2G illustrates determining of mRNAs in a cell, in some embodiments of the invention;

Figs. 3A-3B illustrate the determination of nucleic acids, in accordance with various embodiments of the invention;

Figs. 4A-4B is a non-limiting example of multiple read sequences distributed in a population of different nucleic acid probes, in accordance with certain embodiments of the invention;

Figs. 5A-5E illustrate the determination of nucleic acids, in accordance with another embodiment of the invention;

Figs. 6A-6H illustrate simultaneous determination of multiple nucleic acid species in cells, in certain embodiments of the invention;

Figs. 7A-7E show expression noise of genes and co-variation of expression between different genes determined in accordance with some embodiments of the invention;

Figs. 8A-8E illustrate spatial distribution of RNAs in cells determined in accordance with one embodiment of the invention;

Figs. 9A-9C illustrate simultaneous determination of multiple nucleic acid species in cells, in another embodiment of the invention;

Figs. 10A-10B show expression between different genes determined in accordance with yet another embodiment of the invention;

Fig. 11 is a schematic description of combinatorial labeling, in accordance with another embodiment of the invention;

Figs. 12A-12C show schematic descriptions of Hamming distance, in another embodiment of the invention;

Fig. 13 illustrates the production of a library of probes, in still another embodiment of the invention;

Figs. 14A-14B illustrate fluorescent spot determinations, in another embodiment of the invention;

Figs. 15A-15B illustrate error correction facilitates RNA detection, in yet another embodiment of the invention;

Figs. 16A-16E show characterization of misidentification rates and calling rates, in one embodiment of the invention;

Figs. 17A-17D show characterization of misidentification rates and calling rates, in another embodiment of the invention;

Figs. 18A-18C show a comparison of experiments, in accordance with one embodiment of the invention;

Figs. 19A-19D illustrate decoding and error assessment, in another embodiment of the invention; and

Figs. 20A to 20H illustrate encoding of 130 RNA species and 10 code words as misidentification controls as Codebook 1 and Codebook 2, in another embodiment of the invention.

BRIEF DESCRIPTION OF THE SEQUENCES

SEQ ID NO: 1 is:

GTTGGCGACGAAAGCACTGCGATTGGAACCGTCCCAAGCGTTGCGCTTAATGGA TCATCAATTTTGTCTCACTACGACGGTCAATCGCGCTGCATACTTGCGTCGGTCG GACAAACGAGG;

SEQ ID NO: 2 is CGCAACGCTTGGGACGGTTCCAATCGGATC;

SEQ ID NO: 3 is CGAATGCTCTGGCCTCGAACGAACGATAGC;

SEQ ID NO: 4 is ACAAATCCGACCAGATCGGACGATCATGGG;

SEQ ID NO: 5 is CAAGTATGCAGCGCGATTGACCGTCTCGTT;

SEQ ID NO: 6 is TGCGTCGTCTGGCTAGCACGGCACGCAAAT;

SEQ ID NO: 7 is AAGTCGTACGCCGATGCGCAGCAATTCACT;

SEQ ID NO: 8 is CGAAACATCGGCCACGGTCCCGTTGAACTT;

SEQ ID NO: 9 is ACGAATCCACCGTCCAGCGCGTCAAACAGA;

SEQ ID NO: 10 is CGCGAAATCCCCGTAACGAGCGTCCCTTGC;

SEQ ID NO: 11 is GCATGAGTTGCCTGGCGTTGCGACGACTAA;

SEQ ID NO: 12 is CCGTCGTCTCCGGTCCACCGTTGCGCTTAC;

SEQ ID NO: 13 is GGCCAATGGCCCAGGTCCGTCACGCAATTT;

SEQ ID NO: 14 is TTGATCGAATCGGAGCGTAGCGGAATCTGC;

SEQ ID NO: 15 is CGCGCGGATCCGCTTGTCGGGAACGGATAC;

SEQ ID NO: 16 is GCCTCGATTACGACGGATGTAATTCGGCCG;

SEQ ID NO: 17 is GCCCGTATTCCCGCTTGCGAGTAGGGCAAT

SEQ ID NO: 18 is GTTGGTCGGCACTTGGGTGC;

SEQ ID NO: 19 is CGATGCGCCAATTCCGGTTC;

SEQ ID NO: 20 is CGCGGGCTATATGCGAACCG;

SEQ ID NO: 21 is
TAATACGACTCACTATAGGGAAAGCCGGTTCATCCGGTGG;

SEQ ID NO: 22 is
TAATACGACTCACTATAGGGTGATCATCGCTCGCGGGTTG;

SEQ ID NO: 23 is
TAATACGACTCACTATAGGGCGTGGAGGGCATACAACGC;

SEQ ID NO: 24 is CGCAACGCTTGGGACGGTTCCAATCGGATC/3Cy5Sp/;

SEQ ID NO: 25 is CGAATGCTCTGGCCTCGAACGAACGATAGC/3Cy5Sp/;

SEQ ID NO: 26 is ACAAATCCGACCAGATCGGACGATCATGGG/3Cy5Sp/;

SEQ ID NO: 27 is CAAGTATGCAGCGCGATTGACCGTCTCGTT/3Cy5Sp/;

SEQ ID NO: 28 is GCGGGAAGCACGTGGATTAGGGCATCGACC/3Cy5Sp/;

SEQ ID NO: 29 is AAGTCGTACGCCGATGCGCAGCAATTCACT/3Cy5Sp/;

SEQ ID NO: 30 is CGAAACATCGGCCACGGTCCCGTTGAACTT/3Cy5Sp/;

SEQ ID NO: 31 is ACGAATCCACCGTCCAGCGCGTCAAACAGA/3Cy5Sp/;

SEQ ID NO: 32 is CGCGAAATCCCCGTAACGAGCGTCCCTTGC/3Cy5Sp/;

SEQ ID NO: 33 is GCATGAGTTGCCTGGCGTTGCGACGACTAA/3Cy5Sp/;

SEQ ID NO: 34 is CCGTCGTCTCCGGTCCACCGTTGCGCTTAC/3Cy5Sp/;

SEQ ID NO: 35 is GGCCAATGGCCCAGGTCCGTCACGCAATTT/3Cy5Sp/;
SEQ ID NO: 36 is TTGATCGAATCGGAGCGTAGCGGAATCTGC/3Cy5Sp/;
SEQ ID NO: 37 is CGCGCGGATCCGCTTGTCGGGAACGGATAC/3Cy5Sp/;
SEQ ID NO: 38 is GCCTCGATTACGACGGATGTAATTCGGCCG/3Cy5Sp/; and
SEQ ID NO: 39 is GCCCGTATTCCCGCTTGCGAGTAGGGCAAT/3Cy5Sp/

.

DETAILED DESCRIPTION

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. Certain embodiments are directed to determining nucleic acids, such as mRNA, within cells at relatively high resolutions. In some embodiments, a plurality of nucleic acid probes may be applied to a sample, and their binding within the sample determined, e.g., using fluorescence, to determine locations of the nucleic acid probes within the sample. In some embodiments, codewords may be based on the binding of the plurality of nucleic acid probes, and in some cases, the codewords may define an error-correcting code to reduce or prevent misidentification of the nucleic acids. In certain cases, a relatively large number of different targets may be identified using a relatively small number of labels, e.g., by using various combinatorial approaches.

Two example approaches are now discussed. It should be understood, however, that these are presented by way of explanation and not limitation; other aspects and embodiments are discussed in further detail herein. In one example method, primary probes (also called encoding probes) and secondary probes (also called readout probes) are used, where the primary probes encode "codewords" and bind to target nucleic acids in the sample, and the secondary probes are used to read out the codewords from the primary probes. In another example method, a plurality of different primary probes containing codewords are divided into as many separate pools as there are positions in the codewords, such that each primary probe pool corresponds to a certain value in a certain position of the codewords (e.g., a "one" in the first position as in "1001").

The first example is now described with respect to Fig. 3A. As will be discussed in more detail below, in other embodiments, other configurations may be used as well. In this first example, a series of nucleic acid probes are used to determine nucleic acids within a cell or other sample, e.g., qualitatively or quantitatively. For example, nucleic acids may be identified as being present or absent, and/or the numbers or concentrations of certain nucleic acids may be determined within the cell or other sample. In some cases, the positions of the

probes within the cell or other sample may be determined at relatively high resolutions, and in some cases, at resolutions better than the wavelength of visible light.

This example is generally directed to spatially detecting nucleic acids within a cell or other sample, e.g., at relatively high resolutions. For example, the nucleic acids may be mRNAs, or other nucleic acids described herein. In one set of embodiments, the nucleic acids within the cell may be determined by delivering or applying nucleic acid probes to the cell. In some cases, by using combinatorial approaches, a relatively large number of nucleic acids may be determined using a relatively small number of different labels on the nucleic acid probes. Thus, for example, a relatively small number of experiments may be used to determine a relatively large number of nucleic acids in a sample, e.g., due to simultaneous binding of the nucleic acid probes to different nucleic acids in the sample.

In one set of embodiments, a population of primary nucleic acid probes are applied to the cell (or other sample) that is able to bind nucleic acids suspected of being present within the cell. Afterwards, sequentially, secondary nucleic acid probes that can bind to or otherwise interact with some of the primary nucleic acids are added and determined, e.g., using imaging techniques such as fluorescence microscopy (e.g., conventional fluorescence microscopy), STORM (stochastic optical reconstruction microscopy) or other imaging techniques. After imaging, the secondary nucleic acid probes are inactivated or removed, and a different secondary nucleic acid probe is added to the sample. This may be repeated multiple times with multiple different secondary nucleic acid probes. The pattern of binding of the various secondary nucleic acid probes may be used to determine the primary nucleic acid probes at locations within the cell or other sample, which can be used to determine mRNA or other nucleic acids that are present.

For instance, as is shown in Fig. 3A, a population of nucleic acids 10 within a cell (represented here by nucleic acids 11, 12, and 13) may be exposed to a population of primary nucleic acid probes 20, including probes 21 and 22. The primary nucleic acid probes may contain, for instance, a target sequence that can recognize a nucleic acid (e.g., a sequence within nucleic acid 11). Probes 21 and 22 may contain the same or different targeting sequences, which may bind to or hybridize with the same or different nucleic acids. As an example, as is shown in Fig. 3A, probe 21 contains a first targeting sequence 25 which targets the probe to nucleic acid 11, while probe 22 contains a second targeting sequence 26, not identical to the first targeting sequence 25 and which targets the probe to nucleic acid 12. The target sequence may be substantially complementary to at least a portion of a target

nucleic acid, and enough of the target sequence may be present such that specific binding of the nucleic acid probe to the target nucleic acid can occur.

Primary nucleic acid probes 20 may also contain one or more "read" sequences. Two such read sequences are used in this example, although in other embodiments, there may be one, three, four, or more read sequences present within a primary nucleic acid probe. The read sequences may all independently be the same or different. In addition, in one set of embodiments, different nucleic acid probes may use one or more common read sequences. For example, more than one read sequence may be combinatorially combined on different nucleic acid probes, thereby producing a relatively large number of different nucleic acid probes that can be separately identified, even though only a relatively small number of read sequences are used. Thus, for example, in Fig. 3A, probe 21 contains read sequences 27 and 29, while probe 22 contains read sequences 27 and 28, where the two read sequences 27 are identical, and different from read sequences 28 and 29.

After primary nucleic acid probes 20 have been introduced to the sample and allowed to interact with nucleic acids 11, 12, and 13, one or more secondary nucleic acid probes 30 may be applied to the sample to determine the primary nucleic acid probes. The secondary nucleic acid probes may contain a recognition sequence able to recognize one of the read sequences present within the population of primary nucleic acid probes. For instance, the recognition sequence may be substantially complementary to at least a portion of the read sequence, such that the secondary nucleic acid probe is able to bind to or hybridize with corresponding primary nucleic acid probe. For instance, in this example, recognition sequence 35 is able to recognize read sequence 27. In addition, the secondary nucleic acid probes may contain one or more signaling entities 33. For example, a signaling entity may be a fluorescent entity attached to the probe, or a certain sequence of nucleic acids that can be determined in some fashion. More than one secondary sequence may be used, e.g., sequentially. For example, as shown in this figure, the initial secondary probe 30 may be removed (e.g., as discussed below) and a new secondary probe 31 may be added, containing recognition sequence 36 able to recognize read sequence 28 and one or more signaling entities 33. This may also be repeated multiple times, e.g., to determine read sequence 29 or other read sequences that may be present.

The location of the secondary nucleic acid probes 30, 31, etc. may be determined by determining signaling entity 33. For example, if the signaling entity is fluorescent, then fluorescence microscopy can be used to determine the signaling entity. In some embodiments, imaging of a sample to determine the signaling entity may be used at relatively

high resolutions, and in some cases, super-resolution imaging techniques (e.g., resolutions better than the wavelength of visible light or the diffraction limit of light) may be used. Examples of super-resolution imaging techniques include STORM, or other techniques as discussed herein. In some cases, e.g., with certain super-resolution imaging techniques such as STORM, more than one image of the sample may be acquired.

More than one type of secondary nucleic acid probe may be applied to a cell or other sample. For example, a first secondary nucleic acid probe may be applied that can recognize a first read sequence, then it or its attached signaling entity may be inactivated or removed, and a second secondary nucleic acid probe may be applied that can recognize a second read sequence. This process may be repeated multiple times, each with a different secondary nucleic acid probe, e.g., to determine the read sequences that were present in the various primary nucleic acid probes. Thus, primary nucleic acids within the sample can be determined on the basis of the binding pattern of secondary nucleic acid probes.

For example, a first location within the cell or other sample may exhibit binding of a first secondary probe and a third secondary probe, but not the binding of a second or a fourth secondary probe, while a second location may exhibit a different pattern of binding of various secondary probes. The primary nucleic acid probe that the secondary probes are able to bind to or hybridize with may be determined by considering the pattern of binding of various secondary probes. For instance, referring to Fig. 3A, if a first secondary probe is able to determine read sequence 27, a second secondary probe is able to determine read sequence 28, and a third secondary probe is able to determine read sequence 29, then primary nucleic acid 25 may be determined through the binding of the first and third secondary probes (but not the second secondary probe), while primary nucleic acid 26 may be determined through the binding of the first and second secondary probes (but not the third secondary probe). Similarly, if it is known that first probe 21 contains target sequence 25 while second probe 22 contains target sequence 26, then nucleic acids 11 and 12 may also be determined within the sample, e.g., spatially, based on the binding pattern of the various secondary nucleic acid probes. In addition, it should be noted that due to the presence of more than one read sequence on the primary nucleic acid probes, even though first probe 21 and second probe 22 contains a common read sequence (read sequence 27), these probes may be distinguished in the sample due to the different binding patterns of the various secondary nucleic acid probes.

In certain embodiments, this pattern of binding or hybridization of the secondary nucleic acid probes may be converted into a "codeword." In this example, for instance, the codewords are "101" and "110" for first probe 21 and second probe 22, respectively, where a

value of 1 represents binding and a value of 0 represents no binding. The codewords may also have longer lengths in other embodiments; only three probes are shown here for clarity purposes only. A codeword can be directly related to a specific target nucleic acid sequence of the primary nucleic acid probe. Accordingly, different primary nucleic acid probes may match certain codewords, which can then be used to identify the different targets of the primary nucleic acid probes based on the binding patterns of the secondary probes, even if in some cases, there is overlap in the read sequences of different secondary probes, e.g., as was shown in Fig. 3A. However, if no binding is evident (e.g., for nucleic acid 13), then the codeword would be "000" in this example.

The values in each codeword can also be assigned in different fashions in some embodiments. For example, a value of 0 could represent binding while a value of 1 represents no binding. Similarly, a value of 1 could represent binding of a secondary nucleic acid probe with one type of signaling entity while a value of 0 could represent binding of a secondary nucleic acid probe with another type of distinguishable signaling entity. These signaling entities could be distinguished, for example, via different colors of fluorescence. In some cases, values in codewords need not be confined to 0 and 1. The values could also be drawn from larger alphabets, such as ternary (e.g., 0, 1, and 2) or quaternary (e.g., 0, 1, 2, and 3) systems. Each different value could, for example, be represented by a different distinguishable signaling entity, including (in some cases) one value that may be represented by the absence of signal.

The codewords for each target may be assigned sequentially, or may be assigned at random. For instance, referring to Fig. 3A, a first nucleic acid target may be assigned to 101, while a second nucleic acid target may be assigned to 110. In addition, in some embodiments, the codewords may be assigned using an error-detection system or an error-correcting system, such as a Hamming system, a Golay code, or an extended Hamming system (or a SECDED system, i.e., single error correction, double error detection). Generally speaking, such systems can be used to identify where errors have occurred, and in some cases, such systems can also be used to correct the errors and determine what the correct codeword should have been. For example, a codeword such as 001 may be detected as invalid and corrected using such a system to 101, e.g., if 001 is not previously assigned to a different target sequence. A variety of different error-correcting codes can be used, many of which have previously been developed for use within the computer industry; however, such error-correcting systems have not typically been used within biological systems. Additional examples of such error-correcting codes are discussed in more detail below.

It should also be understood that all possible codewords in a code need not be used in some cases. For example, in some embodiments, codewords that are not used can serve as negative controls. Similarly, in some embodiments, some codewords can be left out because they are more prone to errors in measurement than other codewords. For example, in some implementations, reading a codeword with more values of '1' might be more error-prone that reading a codeword with fewer values of '1.'

It should be understood that the above description is an example of one embodiment of the invention, and that primary and secondary nucleic acid probes are not necessary in all embodiments. For example, in some embodiments, a series of nucleic acid probes containing signaling entities are used to determine nucleic acids within a cell or other sample, without necessarily requiring secondary probes.

For example, turning now to Fig. 3B, nucleic acids 11, 12, and 13 are exposed to different rounds of probes 21, 22, 23, 24, etc. in this example. These probes may each contain a target sequence that can recognize a nucleic acid (e.g., a sequence within nucleic acid 11 or 12). These probes may each target the same nucleic acid, but different regions of the nucleic acid. In addition, some or all of the probes may contain one or more signaling entities, e.g., signaling entity 29 on probe 21. For example, the signaling entity may be a fluorescent entity attached to the probe, or a certain sequence of nucleic acids that can be determined in some fashion.

The first round of probes (e.g. probe 21 and probe 22) may be applied to the cell or other sample. Probe 21 may be allowed to bind to nucleic acid 11 via target sequence 25. Such binding can be determined by determining signaling entity 29. For example, if the signaling entity is fluorescent, then fluorescence microscopy can be used to determine the signaling entity, e.g., spatially within the cell or other sample. In some but not all embodiments, imaging of a sample to determine the signaling entity may be used at relatively high resolutions, and in some cases, super-resolution imaging techniques may be used. Other, different probes may be present as well; for instance, probe 22 containing target sequence 26 may bind to nucleic acid 12, and be determined via signaling entity 29 within probe 22. These may occur, e.g., sequentially or simultaneously. Optionally, probes 21 and 22 may also be removed or inactivated, e.g., between application of different rounds of probes.

Next, a second round of probes (e.g., probe 23) is applied to the sample. In this example, probe 23 is able to bind to nucleic acid 11 via a targeting region, although there is no probe in the second round that is able to bind to nucleic acid 12. Binding of the probes is

allowed to occur as discussed above, and determination of binding may occur via signaling entities. These signaling entities may be the same or different as from the first round of probes. This process may be repeated any number of times with different probes. For example, as is shown in Fig. 3B, round 2 contains probes able to bind to nucleic acid 11, while round 3 contains probes able to bind to nucleic acid 12.

In certain embodiments, each round of binding or hybridization of nucleic acid probes may be converted into a "codeword." In this example, using probes 21, 22, 23, and 24, the codewords 101 or 110 could be formed, where 1 represents binding and 0 represents no binding and the first position corresponds to the binding of probes 21 or 22 while the second position corresponds to the binding of probes 22, and the third position corresponds to the binding of probe 24. A codeword of 000 would represent no binding, e.g., as shown with nucleic acid 13 in this example. A codeword can be directly related to a specific target nucleic acid sequence of the nucleic acid probes, by designing appropriate nucleic acid probes. Thus, for example, 110 may correspond to a first target nucleic acid 12 (e.g., the first and second round of nucleic acid probes containing probes able to target nucleic acid 11, and these probes may target the same or different regions of nucleic acid 11) while 101 may correspond to a second target nucleic acid (e.g., the first and third round of nucleic acid probes containing probes able to target nucleic acid 12, and these probes may target the same or different regions of nucleic acid 12). In addition, it should be noted that each round of probes may contain the same, or different signaling entities as other probes in the same round, and/or other probes in different rounds. For instance, in one set of embodiments, only one signaling entity is used in all of the rounds of probes.

Similar to the above, the codewords for each target may be assigned sequentially, or may be assigned at random. The codewords may be assigned within a code space in some embodiments using an error-detection or an error-correcting system, such as a Hamming system, a Golay code, or an extended Hamming system or a SECDED system (single error correction, double error detection). Generally speaking, such error-correction systems can be used to identify where errors have occurred, and in some cases, such systems can also be used to correct the errors and determine what the correct codeword should have been.

Similar to the above, the values at each position in the codeword can be arbitrarily assigned in certain embodiments to binding or non-binding of probes that contain more than one distinguishable signaling entity.

In some cases, the nucleic acid probes may be formed into "pools" or groups of nucleic acids that share a common feature. For example, probes to all targets with codewords

that contain a 1 in the first position, e.g. 110 and 101 but not 011, may comprise one pool while probes to all targets that contain a 1 in the second position, e.g. 110 and 011 but not 101, may comprise another pool. See also Fig. 1C. In some cases, a nucleic acid probe may be a member of more than one group or pool. Members of a nucleic acid pool may also contain features in addition to target sequences, read sequences, and or signaling entities that allow them to be distinguished from other groups. These features may be short nucleic acid sequences that are used for the amplification, production, or separation of these sequences. The nucleic acid probes of each group may be applied to a sample, e.g., sequentially, as discussed herein.

Thus, in some aspects, the present invention is generally directed to systems and methods for determining nucleic acids within a cell or other sample. The sample may include a cell culture, a suspension of cells, a biological tissue, a biopsy, an organism, or the like. The sample may also be cell-free but nevertheless contain nucleic acids. If the sample contains a cell, the cell may be a human cell, or any other suitable cell, e.g., a mammalian cell, a fish cell, an insect cell, a plant cell, or the like. More than one cell may be present in some cases.

The nucleic acids to be determined may be, for example, DNA, RNA, or other nucleic acids that are present within a cell (or other sample). The nucleic acids may be endogenous to the cell, or added to the cell. For instance, the nucleic acid may be viral, or artificially created. In some cases, the nucleic acid to be determined may be expressed by the cell. The nucleic acid is RNA in some embodiments. The RNA may be coding and/or non-coding RNA. Non-limiting examples of RNA that may be studied within the cell include mRNA, siRNA, rRNA, miRNA, tRNA, lncRNA, snoRNAs, snRNAs, exRNAs, piRNAs, or the like.

In some cases, a significant portion of the nucleic acid within the cell may be studied. For instance, in some cases, enough of the RNA present within a cell may be determined so as to produce a partial or complete transcriptome of the cell. In some cases, at least 4 types of mRNAs are determined within a cell, and in some cases, at least 3, at least 4, at least 7, at least 8, at least 12, at least 14, at least 15, at least 16, at least 22, at least 30, at least 31, at least 32, at least 50, at least 63, at least 64, at least 72, at least 75, at least 100, at least 127, at least 128, at least 140, at least 255, at least 256, at least 500, at least 1,000, at least 1,500, at least 2,000, at least 2,500, at least 3,000, at least 4,000, at least 5,000, at least 7,500, at least 10,000, at least 12,000, at least 15,000, at least 20,000, at least 25,000, at least 30,000, at least 40,000, at least 50,000, at least 75,000, or at least 100,000 types of mRNAs may be determined within a cell.

In some cases, the transcriptome of a cell may be determined. It should be understood that the transriptome generally encompasses all RNA molecules produced within a cell, not just mRNA. Thus, for instance, the transcriptome may also include rRNA, tRNA, siRNA, etc. In some embodiments, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or 100% of the transcriptome of a cell may be determined.

The determination of one or more nucleic acids within the cell or other sample may be qualitative and/or quantitative. In addition, the determination may also be spatial, e.g., the position of the nucleic acid within the cell or other sample may be determined in two or three dimensions. In some embodiments, the positions, number, and/or concentrations of nucleic acids within the cell (or other sample) may be determined.

In some cases, a significant portion of the genome of a cell may be determined. The determined genomic segments may be continuous or interspersed on the genome. For example, in some cases, at least 4 genomic segments are determined within a cell, and in some cases, at least 3, at least 4, at least 7, at least 8, at least 12, at least 14, at least 15, at least 16, at least 22, at least 30, at least 31, at least 32, at least 50, at least 63, at least 64, at least 72, at least 75, at least 100, at least 127, at least 128, at least 140, at least 255, at least 256, at least 500, at least 1,000, at least 1,500, at least 2,000, at least 2,500, at least 3,000, at least 4,000, at least 5,000, at least 7,500, at least 10,000, at least 12,000, at least 15,000, at least 20,000, at least 25,000, at least 30,000, at least 40,000, at least 50,000, at least 75,000, or at least 100,000 genomic segments may be determined within a cell.

In some cases, the entire genome of a cell may be determined. It should be understood that the genome generally encompasses all DNA molecules produced within a cell, not just chromosome DNA. Thus, for instance, the genome may also include, in some cases, mitochondria DNA, chloroplast DNA, plasmid DNA, etc. In some embodiments, at least about 5%, at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or 100% of the genome of a cell may be determined.

As discussed herein, a variety of nucleic acid probes may be used to determine one or more nucleic acids within a cell or other sample. The probes may comprise nucleic acids (or entities that can hybridize to a nucleic acid, e.g., specifically) such as DNA, RNA, LNA (locked nucleic acids), PNA (peptide nucleic acids), or combinations thereof. In some cases,

additional components may also be present within the nucleic acid probes, e.g., as discussed below. Any suitable method may be used to introduce nucleic acid probes into a cell.

For example, in some embodiments, the cell is fixed prior to introducing the nucleic acid probes, e.g., to preserve the positions of the nucleic acids within the cell. Techniques for fixing cells are known to those of ordinary skill in the art. As non-limiting examples, a cell may be fixed using chemicals such as formaldehyde, paraformaldehyde, glutaraldehyde, ethanol, methanol, acetone, acetic acid, or the like. In one embodiment, a cell may be fixed using Hepes-glutamic acid buffer-mediated organic solvent (HOPE).

The nucleic acid probes may be introduced into the cell (or other sample) using any suitable method. In some cases, the cell may be sufficiently permeabilized such that the nucleic acid probes may be introduced into the cell by flowing a fluid containing the nucleic acid probes around the cells. In some cases, the cells may be sufficiently permeabilized as part of a fixation process; in other embodiments, cells may be permeabilized by exposure to certain chemicals such as ethanol, methanol, Triton, or the like. In addition, in some embodiments, techniques such as electroporation or microinjection may be used to introduce nucleic acid probes into a cell or other sample.

Certain aspects of the present invention are generally directed to nucleic acid probes that are introduced into a cell (or other sample). The probes may comprise any of a variety of entities that can hybridize to a nucleic acid, typically by Watson-Crick base pairing, such as DNA, RNA, LNA, PNA, etc., depending on the application. The nucleic acid probe typically contains a target sequence that is able to bind to at least a portion of a target nucleic acid, in some cases specifically. When introduced into a cell or other system, the target system may be able to bind to a specific target nucleic acid (e.g., an mRNA, or other nucleic acids as discussed herein). In some cases, the nucleic acid probes may be determined using signaling entities (e.g., as discussed below), and/or by using secondary nucleic acid probes able to bind to the nucleic acid probes (i.e., to primary nucleic acid probes). The determination of such nucleic acid probes is discussed in detail below.

In some cases, more than one type of (primary) nucleic acid probe may be applied to a sample, e.g., simultaneously. For example, there may be at least 2, at least 5, at least 10, at least 25, at least 50, at least 75, at least 100, at least 300, at least 1,000, at least 3,000, at least 10,000, or at least 30,000 distinguishable nucleic acid probes that are applied to a sample, e.g., simultaneously or sequentially.

The target sequence may be positioned anywhere within the nucleic acid probe (or primary nucleic acid probe or encoding nucleic acid probe). The target sequence may contain

a region that is substantially complementary to a portion of a target nucleic acid. In some cases, the portions may be at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% complementary. In some cases, the target sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 50, at least 60, at least 65, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides in length. In some cases, the target sequence may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the target sequence may have a length of between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc. Typically, complementarity is determined on the basis of Watson-Crick nucleotide base pairing.

The target sequence of a (primary) nucleic acid probe may be determined with reference to a target nucleic acid suspected of being present within a cell or other sample. For example, a target nucleic acid to a protein may be determined using the protein's sequence, by determining the nucleic acids that are expressed to form the protein. In some cases, only a portion of the nucleic acids encoding the protein are used, e.g., having the lengths as discussed above. In addition, in some cases, more than one target sequence that can be used to identify a particular target may be used. For instance, multiple probes can be used, sequentially and/or simultaneously, that can bind to or hybridize to different regions of the same target. Hybridization typically refers to an annealing process by which complementary single-stranded nucleic acids associate through Watson-Crick nucleotide base pairing (e.g., hydrogen bonding, guanine-cytosine and adenine-thymine) to form double-stranded nucleic acid.

In some embodiments, a nucleic acid probe, such as a primary nucleic acid probe, may also comprise one or more "read" sequences. However, it should be understood that read sequences are not necessary in all cases. In some embodiments, the nucleic acid probe may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 or more, 20 or more, 32 or more, 40 or more, 50 or more, 64 or more, 75 or more, 100 or more, 128 or more read

sequences. The read sequences may be positioned anywhere within the nucleic acid probe. If more than one read sequence is present, the read sequences may be positioned next to each other, and/or interspersed with other sequences.

The read sequences, if present, may be of any length. If more than one read sequence is used, the read sequences may independently have the same or different lengths. For instance, the read sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 50, at least 60, at least 65, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides in length. In some cases, the read sequence may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the read sequence may have a length of between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc.

The read sequence may be arbitrary or random in some embodiments. In certain cases, the read sequences are chosen so as to reduce or minimize homology with other components of the cell or other sample, e.g., such that the read sequences do not themselves bind to or hybridize with other nucleic acids suspected of being within the cell or other sample. In some cases, the homology may be less than 10%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, or less than 1%. In some cases, there may be a homology of less than 20 basepairs, less than 18 basepairs, less than 15 basepairs, less than 14 basepairs, less than 13 basepairs, less than 12 basepairs, less than 11 basepairs, or less than 10 basepairs. In some cases, the basepairs are sequential.

In one set of embodiments, a population of nucleic acid probes may contain a certain number of read sequences, which may be less than the number of targets of the nucleic acid probes in some cases. Those of ordinary skill in the art will be aware that if there is one signaling entity and n read sequences, then in general 2^n-1 different nucleic acid targets may be uniquely identified. However, not all possible combinations need be used. For instance, a population of nucleic acid probes may target 12 different nucleic acid sequences, yet contain no more than 8 read sequences. As another example, a population of nucleic acids may target 140 different nucleic acid species, yet contain no more than 16 read sequences. Different

nucleic acid sequence targets may be separately identified by using different combinations of read sequences within each probe. For instance, each probe may contain 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, etc. or more read sequences. In some cases, a population of nucleic acid probes may each contain the same number of read sequences, although in other cases, there may be different numbers of read sequences present on the various probes.

As a non-limiting example, a first nucleic acid probe may contain a first target sequence, a first read sequence, and a second read sequence, while a second, different nucleic acid probe may contain a second target sequence, the same first read sequence, but a third read sequence instead of the second read sequence. Such probes may thereby be distinguished by determining the various read sequences present or associated with a given probe or location, as discussed herein.

In addition, the nucleic acid probes (and their corresponding, complimentary sites on the encoding probes), in certain embodiments, may be made using only 2 or only 3 of the 4 bases, such as leaving out all the "G"s or leaving out all of the "C"s within the probe. Sequences lacking either "G"s or "C"s may form very little secondary structure in certain embodiments, and can contribute to more uniform, faster hybridization.

In some embodiments, the nucleic acid probe may contain a signaling entity. It should be understood that signaling entities are not required in all cases, however; for instance, the nucleic acid probe may be determined using secondary nucleic acid probes in some embodiments, as is discussed in additional detail below. Examples of signaling entities that can be used are also discussed in more detail below.

Other components may also be present within a nucleic acid probe as well. For example, in one set of embodiments, one or more primer sequences may be present, e.g., to allow for enzymatic amplification of probes. Those of ordinary skill in the art will be aware of primer sequences suitable for applications such as amplification (e.g., using PCR or other suitable techniques). Many such primer sequences are available commercially. Other examples of sequences that may be present within a primary nucleic acid probe include, but are not limited to promoter sequences, operons, identification sequences, nonsense sequences, or the like.

Typically, a primer is a single-stranded or partially double-stranded nucleic acid (e.g., DNA) that serves as a starting point for nucleic acid synthesis, allowing polymerase enzymes such as nucleic acid polymerase to extend the primer and replicate the complementary strand. A primer is (e.g., is designed to be) complementary to and to hybridize to a target nucleic acid. In some embodiments, a primer is a synthetic primer. In some embodiments, a primer

is a non-naturally-occurring primer. A primer typically has a length of 10 to 50 nucleotides. For example, a primer may have a length of 10 to 40, 10 to 30, 10 to 20, 25 to 50, 15 to 40, 15 to 30, 20 to 50, 20 to 40, or 20 to 30 nucleotides. In some embodiments, a primer has a length of 18 to 24 nucleotides.

In addition, the components of the nucleic acid probe may be arranged in any suitable order. For instance, in one embodiment, the components may be arranged in a nucleic acid probe as: primer—read sequences—targeting sequence—read sequences—reverse primer. The "read sequences" in this structure may each contain any number (including 0) of read sequences, so long as at least one read sequence is present in the probe. Non-limiting example structures include primer—targeting sequence—read sequences—reverse primer, primer—read sequences—targeting sequence—reverse primer, targeting sequence—primer—targeting sequence—read sequences—reverse primer, targeting sequence—primer—read sequences—targeting sequence—reverse primer, primer—target sequence—read sequences—targeting sequence—reverse primer, targeting sequence—primer—read sequence—reverse primer, targeting sequence—read sequence—primer, read sequence—targeting sequence—primer, read sequence—primer—targeting sequence—reverse primer, etc. In addition, the reverse primer is optional in some embodiments, including in all of the above-described examples.

After introduction of the nucleic acid probes into a cell or other sample, the nucleic acid probes may be directly determined by determining signaling entities (if present), and/or the nucleic acid probes may be determined by using one or more secondary nucleic acid probes, in accordance with certain aspects of the invention. As mentioned, in some cases, the determination may be spatial, e.g., in two or three dimensions. In addition, in some cases, the determination may be quantitative, e.g., the amount or concentration of a primary nucleic acid probe (and of a target nucleic acid) may be determined. Additionally, the secondary probes may comprise any of a variety of entities able to hybridize a nucleic acid, e.g., DNA, RNA, LNA, and/or PNA, etc., depending on the application. Signaling entities are discussed in more detail below.

A secondary nucleic acid probe may contain a recognition sequence able to bind to or hybridize with a read sequence of a primary nucleic acid probe. In some cases, the binding is specific, or the binding may be such that a recognition sequence preferentially binds to or hybridizes with only one of the read sequences that are present. The secondary nucleic acid probe may also contain one or more signaling entities. If more than one secondary nucleic acid probe is used, the signaling entities may be the same or different.

The recognition sequences may be of any length, and multiple recognition sequences may be of the same or different lengths. If more than one recognition sequence is used, the recognition sequences may independently have the same or different lengths. For instance, the recognition sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, or at least 50 nucleotides in length. In some cases, the recognition sequence may be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the recognition sequence may have a length of between 10 and 30, between 20 and 40, or between 25 and 35 nucleotides, etc. In one embodiment, the recognition sequence is of the same length as the read sequence. In addition, in some cases, the recognition sequence may be at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 100% complementary to a read sequence of the primary nucleic acid probe.

As mentioned, in some cases, the secondary nucleic acid probe may comprise one or more signaling entities. Examples of signaling entities are discussed in more detail below.

As discussed, in certain aspects of the invention, nucleic acid probes are used that contain various "read sequences." For example, a population of primary nucleic acid probes may contain certain "read sequences" which can bind certain of the secondary nucleic acid probes, and the locations of the primary nucleic acid probes are determined within the sample using secondary nucleic acid probes, e.g., which comprise a signaling entity. As mentioned, in some cases, a population of read sequences may be combined in various combinations to produce different nucleic acid probes, e.g., such that a relatively small number of read sequences may be used to produce a relatively large number of different nucleic acid probes.

Thus, in some cases, a population of primary nucleic acid probes (or other nucleic acid probes) may each contain a certain number of read sequences, some of which are shared between different primary nucleic acid probes such that the total population of primary nucleic acid probes may contain a certain number of read sequences. A population of nucleic acid probes may have any suitable number of read sequences. For example, a population of primary nucleic acid probes may have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 etc. read sequences. More than 20 are also possible in some embodiments. In addition, in some cases, a population of nucleic acid probes may, in total, have 1 or more, 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or

more, 11 or more, 12 or more, 13 or more, 14 or more, 15 or more, 16 or more, 20 or more, 24 or more, 32 or more, 40 or more, 50 or more, 60 or more, 64 or more, 100 or more, 128 or more, etc. of possible read sequences present, although some or all of the probes may each contain more than one read sequence, as discussed herein. In addition, in some embodiments, the population of nucleic acid probes may have no more than 100, no more than 80, no more than 64, no more than 60, no more than 50, no more than 40, no more than 32, no more than 24, no more than 20, no more than 16, no more than 15, no more than 14, no more than 13, no more than 12, no more than 11, no more than 10, no more than 9, no more than 8, no more than 7, no more than 6, no more than 5, no more than 4, no more than 3, or no more than two read sequences present. Combinations of any of these are also possible, e.g., a population of nucleic acid probes may comprise between 10 and 15 read sequences in total.

As a non-limiting example of an approach to combinatorially producing a relatively large number of nucleic acid probes from a relatively small number of read sequences, in a population of 6 different types of nucleic acid probes, each comprising one or more read sequences, the total number of read sequences within the population may be no greater than 4. It should be understood that although 4 read sequences are used in this example for ease of explanation, in other embodiments, larger numbers of nucleic acid probes may be realized, for example, using 5, 8, 10, 16, 32, etc. or more read sequences, or any other suitable number of read sequences described herein, depending on the application. Referring now to Fig. 4A, if each of the primary nucleic acid probes contains two different read sequences, then by using 4 such read sequences (A, B, C, and D), up to 6 probes may be separately identified. It should be noted that in this example, the ordering of read sequences on a nucleic acid probe is not essential, i.e., "AB" and "BA" may be treated as being synonymous (although in other embodiments, the ordering of read sequences may be essential and "AB" and "BA" may not necessarily be synonymous). Similarly, if 5 read sequences are used (A, B, C, D, and E) in the population of primary nucleic acid probes, up to 10 probes may be separately identified, as is shown in Fig. 4B. For example, one of ordinary skill in the art would understand that, for k read sequences in a population with n read sequences on each probe, up to $\binom{n}{k}$ different probes may be produced, assuming that the ordering of read sequences is not essential; because not all of the probes need to have the same number of read sequences and not all combinations of read sequences need to be used in every embodiment, either more or less than this number of different probes may also be used in certain embodiments. In addition, it should also be understood that the number of read sequences on each probe need not be

identical in some embodiments. For instance example, some probes may contain 2 read sequences while other probes may contain 3 read sequences.

In some aspects, the read sequences and/or the pattern of binding of nucleic acid probes within a sample may be used to define an error-detecting and/or an error-correcting code, for example, to reduce or prevent misidentification or errors of the nucleic acids, e.g., as was discussed with reference to Fig. 3. Thus, for example, if binding is indicated (e.g., as determined using a signaling entity), then the location may be identified with a "1"; conversely, if no binding is indicated, then the location may be identified with a "0" (or vice versa, in some cases). Multiple rounds of binding determinations, e.g., using different nucleic acid probes, can then be used to create a "codeword," e.g., for that spatial location. In some embodiments, the codeword may be subjected to error detection and/or correction. For instance, the codewords may be organized such that, if no match is found for a given set of read sequences or binding pattern of nucleic acid probes, then the match may be identified as an error, and optionally, error correction may be applied sequences to determine the correct target for the nucleic acid probes. In some cases, the codewords may have fewer "letters" or positions that the total number of nucleic acids encoded by the codewords, e.g. where each codeword encodes a different nucleic acid.

Such error-detecting and/or the error-correction code may take a variety of forms. A variety of such codes have previously been developed in other contexts such as the telecommunications industry, such as Golay codes or Hamming codes. In one set of embodiments, the read sequences or binding patterns of the nucleic acid probes are assigned such that not every possible combination is assigned.

For example, if 4 read sequences are possible and a primary nucleic acid probe contains 2 read sequences, then up to 6 primary nucleic acid probes could be identified; but the number of primary nucleic acid probes used may be less than 6. Similarly, for k read sequences in a population with n read sequences on each primary nucleic acid probe, $\binom{n}{k}$ different probes may be produced, but the number of primary nucleic acid probes that are used may be any number more or less than $\binom{n}{k}$. In addition, these may be randomly assigned, or assigned in specific ways to increase the ability to detect and/or correct errors.

As another example, if multiple rounds of nucleic acid probes are used, the number of rounds may be arbitrarily chosen. If in each round, each target can give two possible outcomes, such as being detected or not being detected, up to 2^n different targets may be possible for n rounds of probes, but the number of nucleic acid targets that are actually used

may be any number less than 2^n. For example, if in each round, each target can give more than two possible outcomes, such as being detected in different color channels, more than 2^n (e.g. 3^n, 4^n ...) different targets may be possible for n rounds of probes. In some cases, the number of nucleic acid targets that are actually used may be any number less than this number. In addition, these may be randomly assigned, or assigned in specific ways to increase the ability to detect and/or correct errors.

For example, in one set of embodiments, the codewords or nucleic acid probes may be assigned within a code space such that the assignments are separated by a Hamming distance, which measures the number of incorrect "reads" in a given pattern that cause the nucleic acid probe to be misinterpreted as a different valid nucleic acid probe. In certain cases, the Hamming distance may be at least 2, at least 3, at least 4, at least 5, at least 6, or the like. In addition, in one set of embodiments, the assignments may be formed as a Hamming code, for instance, a Hamming(7, 4) code, a Hamming(15, 11) code, a Hamming(31, 26) code, a Hamming(63, 57) code, a Hamming(127, 120) code, etc. In another set of embodiments, the assignments may form a SECDED code, e.g., a SECDED(8,4) code, a SECDED(16,4) code, a SCEDED(16, 11) code, a SCEDED(22, 16) code, a SCEDED(39, 32) code, a SCEDED(72, 64) code, etc. In yet another set of embodiments, the assignments may form an extended binary Golay code, a perfect binary Golay code, or a ternary Golay code. In another set of embodiments, the assignments may represent a subset of the possible values taken from any of the codes described above.

For example, a code with the same error correcting properties of the SECDED code may be formed by using only binary words that contain a fixed number of '1' bits, such as 4, to encode the targets. In another set of embodiments, the assignments may represent a subset of the possible values taken from codes described above for the purpose of addressing asymmetric readout errors. For example, in some cases, a code in which the number of '1' bits may be fixed for all used binary words may eliminate the biased measurement of words with different numbers of '1's when the rate at which '0' bits are measured as '1's or '1' bits are measured as '0's are different.

Accordingly, in some embodiments, once the codeword is determined (e.g., as discussed herein), the codeword may be compared to the known nucleic acid codewords. If a match is found, then the nucleic acid target can be identified or determined. If no match is found, then an error in the reading of the codeword may be identified. In some cases, error correction can also be applied to determine the correct codeword, and thus resulting in the correct identity of the nucleic acid target. In some cases, the codewords may be selected such

that, assuming that there is only one error present, only one possible correct codeword is available, and thus, only one correct identity of the nucleic acid target is possible. In some cases, this may also be generalized to larger codeword spacings or Hamming distances; for instance, the codewords may be selected such that if two, three, or four errors are present (or more in some cases), only one possible correct codeword is available, and thus, only one correct identity of the nucleic acid targets is possible.

The error-correcting code may be a binary error-correcting code, or it may be based on other numbering systems, e.g., ternary or quaternary error-correcting codes. For instance, in one set of embodiments, more than one type of signaling entity may be used and assigned to different numbers within the error-correcting code. Thus, as a non-limiting example, a first signaling entity (or more than one signaling entity, in some cases) may be assigned as "1" and a second signaling entity (or more than one signaling entity, in some cases) may be assigned as "2" (with "0" indicating no signaling entity present), and the codewords distributed to define a ternary error-correcting code. Similarly, a third signaling entity may additionally be assigned as "3" to make a quaternary error-correcting code, etc.

As discussed above, in certain aspects, signaling entities are determined, e.g., to determine nucleic acid probes and/or to create codewords. In some cases, signaling entities within a sample may be determined, e.g., spatially, using a variety of techniques. In some embodiments, the signaling entities may be fluorescent, and techniques for determining fluorescence within a sample, such as fluorescence microscopy or confocal microscopy, may be used to spatially identify the positions of signaling entities within a cell. In some cases, the positions of entities within the sample may be determined in two or even three dimensions. In addition, in some embodiments, more than one signaling entity may be determined at a time (e.g., signaling entities with different colors or emissions), and/or sequentially.

In addition, in some embodiments, a confidence level for the identified nucleic acid target may be determined. For example, the confidence level may be determined using a ratio of the number of exact matches to the number of matches having one or more one-bit errors. In some cases, only matches having a confidence ratio greater than a certain value may be used. For instance, in certain embodiments, matches may be accepted only if the confidence ratio for the match is greater than about 0.01, greater than about 0.03, greater than about 0.05, greater than about 0.1, greater than about 0.3, greater than about 0.5, greater than about 1, greater than about 3, greater than about 5, greater than about 10, greater than about 30, greater than about 50, greater than about 100, greater than about 300, greater than about

500, greater than about 1000, or any other suitable value. In addition, in some embodiments, matches may be accepted only if the confidence ratio for the identified nucleic acid target is greater than an internal standard or false positive control by about 0.01, about 0.03, about 0.05, about 0.1, about 0.3, about 0.5, about 1, about 3, about 5, about 10, about 30, about 50, about 100, about 300, about 500, about 1000, or any other suitable value

In some embodiments, the spatial positions of the entities (and thus, nucleic acid probes that the entities may be associated with) may be determined at relatively high resolutions. For instance, the positions may be determined at spatial resolutions of better than about 100 micrometers, better than about 30 micrometers, better than about 10 micrometers, better than about 3 micrometers, better than about 1 micrometer, better than about 800 nm, better than about 600 nm, better than about 500 nm, better than about 400 nm, better than about 300 nm, better than about 200 nm, better than about 100 nm, better than about 90 nm, better than about 80 nm, better than about 70 nm, better than about 60 nm, better than about 50 nm, better than about 40 nm, better than about 30 nm, better than about 20 nm, or better than about 10 nm, etc.

There are a variety of techniques able to determine or image the spatial positions of entities optically, e.g., using fluorescence microscopy. In some cases, the spatial positions may be determined at super resolutions, or at resolutions better than the wavelength of light or the diffraction limit. Non-limiting examples include STORM (stochastic optical reconstruction microscopy), STED (stimulated emission depletion microscopy), NSOM (Near-field Scanning Optical Microscopy), 4Pi microscopy, SIM (Structured Illumination Microscopy), SMI (Spatially Modulated Illumination) microscopy, RESOLFT (Reversible Saturable Optically Linear Fluorescence Transition Microscopy), GSD (Ground State Depletion Microscopy), SSIM (Saturated Structured-Illumination Microscopy), SPDM (Spectral Precision Distance Microscopy), Photo-Activated Localization Microscopy (PALM), Fluorescence Photoactivation Localization Microscopy (FPALM), LIMON (3D Light Microscopical Nanosizing Microscopy), Super-resolution optical fluctuation imaging (SOFI), or the like. See, e.g., U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*; or Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*, each incorporated herein by reference in their entireties.

As an illustrative non-limiting example, in one set of embodiments, the sample may be imaged with a high numerical aperture, oil immersion objective with 100X magnification and light collected on an electron-multiplying CCD camera. In another example, the sample could be imaged with a high numerical aperture, oil immersion lens with 40X magnification and light collected with a wide-field scientific CMOS camera. With different combinations of objectives and cameras, a single field of view may correspond to no less than 40 x 40 microns, 80 x 80 microns, 120 x 120 microns, 240 x 240 microns, 340 x 340 microns, or 500 x 500 microns, etc. in various non-limiting embodiments. Similarly, a single camera pixel may correspond, in some embodiments, to regions of the sample of no less than 80x80 nm, 120x120 nm, 160x160 nm, 240x240 nm, or 300x300 nm, etc. In another example, the sample may be imaged with a low numerical aperture, air lens with 10X magnification and light collected with a sCMOS camera. In additional embodiments, the sample may be optically sectioned by illuminating it via a single or multiple scanned diffraction limited foci generated either by scanning mirrors or a spinning disk and the collected passed through a single or multiple pinholes. In another embodiment, the sample may also be illuminated via thin sheet of light generated via any one of multiple methods known to those versed in the art.

In one embodiment, the sample may be illuminated by single Gaussian mode laser lines. In some embodiments, the illumination profiled may be flattened by passing these laser lines through a multimode fiber that is vibrated via piezo-electric or other mechanical means. In some embodiments, the illumination profile may be flattened by passing single-mode, Gaussian beams through a variety of refractive beam shapers, such as the piShaper or a series of stacked Powell lenses. In yet another set of embodiments, the Gaussian beams may be passed through a variety of different diffusing elements, such as ground glass or engineered diffusers, which may be spun in some cases at high speeds to remove residual laser speckle. In yet another embodiment, laser illumination may be passed through a series of lenslet arrays to produce overlapping images of the illumination that approximate a flat illumination field.

In some embodiments, the centroids of the spatial positions of the entities may be determined. For example, a centroid of a signaling entity may be determined within an image or series of images using image analysis algorithms known to those of ordinary skill in the art. In some cases, the algorithms may be selected to determine non-overlapping single emitters and/or partially overlapping single emitters in a sample. Non-limiting examples of suitable techniques include a maximum likelihood algorithm, a least squares algorithm, a

Bayesian algorithm, a compressed sensing algorithm, or the like. Combinations of these techniques may also be used in some cases.

In addition, the signaling entity may be inactivated in some cases. For example, in some embodiments, a first secondary nucleic acid probe containing a signaling entity may be applied to a sample that can recognize a first read sequence, then the first secondary nucleic acid probe can be inactivated before a second secondary nucleic acid probe is applied to the sample. If multiple signaling entities are used, the same or different techniques may be used to inactivate the signaling entities, and some or all of the multiple signaling entities may be inactivated, e.g., sequentially or simultaneously.

Inactivation may be caused by removal of the signaling entity (e.g., from the sample, or from the nucleic acid probe, etc.), and/or by chemically altering the signaling entity in some fashion, e.g., by photobleaching the signaling entity, bleaching or chemically altering the structure of the signaling entity, e.g., by reduction, etc.). For instance, in one set of embodiments, a fluorescent signaling entity may be inactivated by chemical or optical techniques such as oxidation, photobleaching, chemically bleaching, stringent washing or enzymatic digestion or reaction by exposure to an enzyme, dissociating the signaling entity from other components (e.g., a probe), chemical reaction of the signaling entity (e.g., to a reactant able to alter the structure of the signaling entity) or the like. For instance, bleaching may occur by exposure to oxygen, reducing agents, or the signaling entity could be chemically cleaved from the nucleic acid probe and washed away via fluid flow.

In some embodiments, various nucleic acid probes (including primary and/or secondary nucleic acid probes) may include one or more signaling entities. If more than one nucleic acid probe is used, the signaling entities may each by the same or different. In certain embodiments, a signaling entity is any entity able to emit light. For instance, in one embodiment, the signaling entity is fluorescent. In other embodiments, the signaling entity may be phosphorescent, radioactive, absorptive, etc. In some cases, the signaling entity is any entity that can be determined within a sample at relatively high resolutions, e.g., at resolutions better than the wavelength of visible light or the diffraction limit. The signaling entity may be, for example, a dye, a small molecule, a peptide or protein, or the like. The signaling entity may be a single molecule in some cases. If multiple secondary nucleic acid probes are used, the nucleic acid probes may comprise the same or different signaling entities.

Non-limiting examples of signaling entities include fluorescent entities (fluorophores) or phosphorescent entities, for example, cyanine dyes (e.g., Cy2, Cy3, Cy3B, Cy5, Cy5.5,

Cy7, etc.), Alexa Fluor dyes, Atto dyes, photoswtichable dyes, photoactivatable dyes, fluorescent dyes, metal nanoparticles, semiconductor nanoparticles or "quantum dots", fluorescent proteins such as GFP (Green Fluorescent Protein), or photoactivabale fluorescent proteins, such as PAGFP, PSCFP, PSCFP2, Dendra, Dendra2, EosFP, tdEos, mEos2, mEos3, PAmCherry, PAtagRFP, mMaple, mMaple2, and mMaple3. Other suitable signaling entities are known to those of ordinary skill in the art. See, e.g., U.S. Pat. No. 7,838,302 or U.S. Pat. Apl. Ser. No. 61/979,436, each incorporated herein by reference in its entirety.

In one set of embodiments, the signaling entity may be attached to an oligonucleotide sequence via a bond that can be cleaved to release the signaling entity. In one set of embodiments, a fluorophore may be conjugated to an oligonucleotide via a cleavable bond, such as a photocleavable bond. Non-limiting examples of photocleavable bonds include, but are not limited to, 1-(2-nitrophenyl)ethyl, 2-nitrobenzyl, biotin phosphoramidite, acrylic phosphoramidite, diethylaminocoumarin, 1-(4,5-dimethoxy-2-nitrophenyl)ethyl, cyclo-dodecyl (dimethoxy-2-nitrophenyl)ethyl, 4-aminomethyl-3-nitrobenzyl, (4-nitro-3-(1-chlorocarbonyloxyethyl)phenyl)methyl-S-acetylthioic acid ester, (4-nitro-3-(1-thlorocarbonyloxyethyl)phenyl)methyl-3-(2-pyridyldithiopropionic acid) ester, 3-(4,4'-dimethoxytrityl)-1-(2-nitrophenyl)-propane-1,3-diol-[2- cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-trifluoroacetylcaproamidomethyl)phenyl]-ethyl-[2-cyano-ethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-(4,4'-dimethoxytrityloxy)butyramidomethyl)phenyl]-ethyl-[2-cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-(N-(4,4'-dimethoxytrityl))-biotinamidocaproamido-methyl)phenyl]-ethyl-[2-cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, or similar linkers. In another set of embodiments, the fluorophore may be conjugated to an oligonucleotide via a disulfide bond. The disulfide bond may be cleaved by a variety of reducing agents such as, but not limited to, dithiothreitol, dithioerythritol, beta-mercaptoethanol, sodium borohydride, thioredoxin, glutaredoxin, trypsinogen, hydrazine, diisobutylaluminum hydride, oxalic acid, formic acid, ascorbic acid, phosphorous acid, tin chloride, glutathione, thioglycolate, 2,3-dimercaptopropanol, 2-mercaptoethylamine, 2-aminoethanol, tris(2-carboxyethyl)phosphine, bis(2-mercaptoethyl) sulfone, N,N'-dimethyl-N,N'-bis(mercaptoacetyl)hydrazine, 3-mercaptoproptionate, dimethylformamide, thiopropyl-agarose, tri-*n*-butylphosphine, cysteine, iron sulfate, sodium sulfite, phosphite, hypophosphite, phosphorothioate, or the like, and/or combinations of any of these. In another embodiment, the fluorophore may be conjugated to an oligonucleotide via one or more phosphorothioate modified nucleotides in which the sulfur modification replaces the bridging and/or non-bridging oxygen. The fluorophore may be

cleaved from the oligonucleotide, in certain embodiments, via addition of compounds such as but not limited to iodoethanol, iodine mixed in ethanol, silver nitrate, or mercury chloride. In yet another set of embodiments, the signaling entity may be chemically inactivated through reduction or oxidation. For example, in one embodiment, a chromophore such as Cy5 or Cy7 may be reduced using sodium borohydride to a stable, non-fluorescence state. In still another set of embodiments, a fluorophore may be conjugated to an oligonucleotide via an azo bond, and the azo bond may be cleaved with 2-[(2-N-arylamino)phenylazo]pyridine. In yet another set of embodiments, a fluorophore may be conjugated to an oligonucleotide via a suitable nucleic acid segment that can be cleaved upon suitable exposure to DNAse, e.g., an exodeoxyribonuclease or an endodeoxyribonuclease. Examples include, but are not limited to, deoxyribonuclease I or deoxyribonuclease II. In one set of embodiments, the cleavage may occur via a restriction endonuclease. Non-limiting examples of potentially suitable restriction endonucleases include BamHI, BsrI, NotI, XmaI, PspAI, DpnI, MboI, MnlI, Eco57I, Ksp632I, DraIII, AhaII, SmaI, MluI, HpaI, ApaI, BclI, BstEII, TaqI, EcoRI, SacI, HindII, HaeII, DraII, Tsp509I, Sau3AI, PacI, etc. Over 3000 restriction enzymes have been studied in detail, and more than 600 of these are available commercially. In yet another set of embodiments, a fluorophore may be conjugated to biotin, and the oligonucleotide conjugated to avidin or streptavidin. An interaction between biotin and avidin or streptavidin allows the fluorophore to be conjugated to the oligonucleotide, while sufficient exposure to an excess of addition, free biotin could "outcompete" the linkage and thereby cause cleavage to occur. In addition, in another set of embodiments, the probes may be removed using corresponding "toe-hold-probes," which comprise the same sequence as the probe, as well as an extra number of bases of homology to the encoding probes (e.g., 1-20 extra bases, for example, 5 extra bases). These probes may remove the labeled readout probe through a strand-displacement interaction.

As used herein, the term "light" generally refers to electromagnetic radiation, having any suitable wavelength (or equivalently, frequency). For instance, in some embodiments, the light may include wavelengths in the optical or visual range (for example, having a wavelength of between about 400 nm and about 700 nm, i.e., "visible light"), infrared wavelengths (for example, having a wavelength of between about 300 micrometers and 700 nm), ultraviolet wavelengths (for example, having a wavelength of between about 400 nm and about 10 nm), or the like. In certain cases, as discussed in detail below, more than one entity may be used, i.e., entities that are chemically different or distinct, for example,

structurally. However, in other cases, the entities may be chemically identical or at least substantially chemically identical.

In one set of embodiments, the signaling entity is "switchable," i.e., the entity can be switched between two or more states, at least one of which emits light having a desired wavelength. In the other state(s), the entity may emit no light, or emit light at a different wavelength. For instance, an entity may be "activated" to a first state able to produce light having a desired wavelength, and "deactivated" to a second state not able to emit light of the same wavelength. An entity is "photoactivatable" if it can be activated by incident light of a suitable wavelength. As a non-limiting example, Cy5, can be switched between a fluorescent and a dark state in a controlled and reversible manner by light of different wavelengths, i.e., 633 nm (or 642nm, 647nm, 656 nm) red light can switch or deactivate Cy5 to a stable dark state, while 405 nm green light can switch or activate the Cy5 back to the fluorescent state. In some cases, the entity can be reversibly switched between the two or more states, e.g., upon exposure to the proper stimuli. For example, a first stimuli (e.g., a first wavelength of light) may be used to activate the switchable entity, while a second stimuli (e.g., a second wavelength of light) may be used to deactivate the switchable entity, for instance, to a non-emitting state. Any suitable method may be used to activate the entity. For example, in one embodiment, incident light of a suitable wavelength may be used to activate the entity to emit light, i.e., the entity is "photoswitchable." Thus, the photoswitchable entity can be switched between different light-emitting or non-emitting states by incident light, e.g., of different wavelengths. The light may be monochromatic (e.g., produced using a laser) or polychromatic. In another embodiment, the entity may be activated upon stimulation by electric field and/or magnetic field. In other embodiments, the entity may be activated upon exposure to a suitable chemical environment, e.g., by adjusting the pH, or inducing a reversible chemical reaction involving the entity, etc. Similarly, any suitable method may be used to deactivate the entity, and the methods of activating and deactivating the entity need not be the same. For instance, the entity may be deactivated upon exposure to incident light of a suitable wavelength, or the entity may be deactivated by waiting a sufficient time.

Typically, a "switchable" entity can be identified by one of ordinary skill in the art by determining conditions under which an entity in a first state can emit light when exposed to an excitation wavelength, switching the entity from the first state to the second state, e.g., upon exposure to light of a switching wavelength, then showing that the entity, while in the second state can no longer emit light (or emits light at a much reduced intensity) when exposed to the excitation wavelength.

In one set of embodiments, as discussed, a switchable entity may be switched upon exposure to light. In some cases, the light used to activate the switchable entity may come from an external source, e.g., a light source such as a laser light source, another light-emitting entity proximate the switchable entity, etc. The second, light emitting entity, in some cases, may be a fluorescent entity, and in certain embodiments, the second, light-emitting entity may itself also be a switchable entity.

In some embodiments, the switchable entity includes a first, light-emitting portion (e.g., a fluorophore), and a second portion that activates or "switches" the first portion. For example, upon exposure to light, the second portion of the switchable entity may activate the first portion, causing the first portion to emit light. Examples of activator portions include, but are not limited to, Alexa Fluor 405 (Invitrogen), Alexa Fluor 488 (Invitrogen), Cy2 (GE Healthcare), Cy3 (GE Healthcare), Cy3B (GE Healthcare), Cy3.5 (GE Healthcare), or other suitable dyes. Examples of light-emitting portions include, but are not limited to, Cy5, Cy5.5 (GE Healthcare), Cy7 (GE Healthcare), Alexa Fluor 647 (Invitrogen), Alexa Fluor 680 (Invitrogen), Alexa Fluor 700 (Invitrogen), Alexa Fluor 750 (Invitrogen), Alexa Fluor 790 (Invitrogen), DiD, DiR, YOYO-3 (Invitrogen), YO-PRO-3 (Invitrogen), TOT-3 (Invitrogen), TO-PRO-3 (Invitrogen) or other suitable dyes. These may linked together, e.g., covalently, for example, directly, or through a linker, e.g., forming compounds such as, but not limited to, Cy5-Alexa Fluor 405, Cy5-Alexa Fluor 488, Cy5-Cy2, Cy5-Cy3, Cy5-Cy3.5, Cy5.5-Alexa Fluor 405, Cy5.5-Alexa Fluor 488, Cy5.5-Cy2, Cy5.5-Cy3, Cy5.5-Cy3.5, Cy7-Alexa Fluor 405, Cy7-Alexa Fluor 488, Cy7-Cy2, Cy7-Cy3, Cy7-Cy3.5, Alexa Fluor 647-Alexa Fluor 405, Alexa Fluor 647-Alexa Fluor 488, Alexa Fluor 647-Cy2, Alexa Fluor 647-Cy3, Alexa Fluor 647-Cy3.5, Alexa Fluor 750-Alexa Fluor 405, Alexa Fluor 750-Alexa Fluor 488, Alexa Fluor 750-Cy2, Alexa Fluor 750-Cy3, or Alexa Fluor 750-Cy3.5. Those of ordinary skill in the art will be aware of the structures of these and other compounds, many of which are available commercially. The portions may be linked via a covalent bond, or by a linker, such as those described in detail below. Other light-emitting or activator portions may include portions having two quaternized nitrogen atoms joined by a polymethine chain, where each nitrogen is independently part of a heteroaromatic moiety, such as pyrrole, imidazole, thiazole, pyridine, quinoine, indole, benzothiazole, etc., or part of a nonaromatic amine. In some cases, there may be 5, 6, 7, 8, 9, or more carbon atoms between the two nitrogen atoms.

In certain cases, the light-emitting portion and the activator portions, when isolated from each other, may each be fluorophores, i.e., entities that can emit light of a certain,

emission wavelength when exposed to a stimulus, for example, an excitation wavelength. However, when a switchable entity is formed that comprises the first fluorophore and the second fluorophore, the first fluorophore forms a first, light-emitting portion and the second fluorophore forms an activator portion that switches that activates or "switches" the first portion in response to a stimulus. For example, the switchable entity may comprise a first fluorophore directly bonded to the second fluorophore, or the first and second entity may be connected via a linker or a common entity. Whether a pair of light-emitting portion and activator portion produces a suitable switchable entity can be tested by methods known to those of ordinary skills in the art. For example, light of various wavelength can be used to stimulate the pair and emission light from the light-emitting portion can be measured to determined wither the pair makes a suitable switch.

As a non-limiting example, Cy3 and Cy5 may be linked together to form such an entity. In this example, Cy3 is an activator portion that is able to activate Cy5, the light-emission portion. Thus, light at or near the absorption maximum (e.g., near 532 nm light for Cy3) of the activation or second portion of the entity may cause that portion to activate the first, light-emitting portion, thereby causing the first portion to emit light (e.g., near 647 nm for Cy5). See, e.g., U.S. Pat. No. 7,838,302, incorporated herein by reference in its entirety. In some cases, the first, light-emitting portion can subsequently be deactivated by any suitable technique (e.g., by directing 647 nm red light to the Cy5 portion of the molecule).

Other non-limiting examples of potentially suitable activator portions include 1,5 IAEDANS, 1,8-ANS, 4-Methylumbelliferone, 5-carboxy-2,7-dichlorofluorescein, 5-Carboxyfluorescein (5-FAM), 5-Carboxynapthofluorescein, 5-Carboxytetramethylrhodamine (5-TAMRA), 5-FAM (5-Carboxyfluorescein), 5-HAT (Hydroxy Tryptamine), 5-Hydroxy Tryptamine (HAT), 5-ROX (carboxy-X-rhodamine), 5-TAMRA (5-Carboxytetramethylrhodamine), 6-Carboxyrhodamine 6G, 6-CR 6G, 6-JOE, 7-Amino-4-methylcoumarin, 7-Aminoactinomycin D (7-AAD), 7-Hydroxy-4-methylcoumarin, 9-Amino-6-chloro-2-methoxyacridine, ABQ, Acid Fuchsin, ACMA (9-Amino-6-chloro-2-methoxyacridine), Acridine Orange, Acridine Red, Acridine Yellow, Acriflavin, Acriflavin Feulgen SITSA, Alexa Fluor 350, Alexa Fluor 405, Alexa Fluor 430, Alexa Fluor 488, Alexa Fluor 500, Alexa Fluor 514, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 555, Alexa Fluor 568, Alexa Fluor 594, Alexa Fluor 610, Alexa Fluor 633, Alexa Fluor 635, Alizarin Complexon, Alizarin Red, AMC, AMCA-S, AMCA (Aminomethylcoumarin), AMCA-X, Aminoactinomycin D, Aminocoumarin, Aminomethylcoumarin (AMCA), Anilin Blue, Anthrocyl stearate, APTRA-BTC, APTS, Astrazon Brilliant Red 4G, Astrazon Orange R,

Astrazon Red 6B, Astrazon Yellow 7 GLL, Atabrine, ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 520, ATTO 532, ATTO 550, ATTO 565, ATTO 590, ATTO 594, ATTO 610, ATTO 611X, ATTO 620, ATTO 633, ATTO 635, ATTO 647, ATTO 647N, ATTO 655, ATTO 680, ATTO 700, ATTO 725, ATTO 740, ATTO-TAG CBQCA, ATTO-TAG FQ, Auramine, Aurophosphine G, Aurophosphine, BAO 9 (Bisaminophenyloxadiazole), BCECF (high pH), BCECF (low pH), Berberine Sulphate, Bimane, Bisbenzamide, Bisbenzimide (Hoechst), bis-BTC, Blancophor FFG, Blancophor SV, BOBO -1, BOBO -3, Bodipy 492/515, Bodipy 493/503, Bodipy 500/510, Bodipy 505/515, Bodipy 530/550, Bodipy 542/563, Bodipy 558/568, Bodipy 564/570, Bodipy 576/589, Bodipy 581/591, Bodipy 630/650-X, Bodipy 650/665-X, Bodipy 665/676, Bodipy Fl, Bodipy FL ATP, Bodipy Fl-Ceramide, Bodipy R6G, Bodipy TMR, Bodipy TMR-X conjugate, Bodipy TMR-X, SE, Bodipy TR, Bodipy TR ATP, Bodipy TR-X SE, BO-PRO -1, BO-PRO -3, Brilliant Sulphoflavin FF, BTC, BTC-5N, Calcein, Calcein Blue, Calcium Crimson, Calcium Green, Calcium Green-1 Ca^{2+} Dye, Calcium Green-2 Ca^{2+}, Calcium Green-5N Ca^{2+}, Calcium Green-C18 Ca^{2+}, Calcium Orange, Calcofluor White, Carboxy-X-rhodamine (5-ROX), Cascade Blue, Cascade Yellow, Catecholamine, CCF2 (GeneBlazer), CFDA, Chromomycin A, Chromomycin A, CL-NERF, CMFDA, Coumarin Phalloidin, CPM Methylcoumarin, CTC, CTC Formazan, Cy2, Cy3.1 8, Cy3.5, Cy3, Cy5.1 8, cyclic AMP Fluorosensor (FiCRhR), Dabcyl, Dansyl, Dansyl Amine, Dansyl Cadaverine, Dansyl Chloride, Dansyl DHPE, Dansyl fluoride, DAPI, Dapoxyl, Dapoxyl 2, Dapoxyl 3' DCFDA, DCFH (Dichlorodihydrofluorescein Diacetate), DDAO, DHR (Dihydorhodamine 123), Di-4-ANEPPS, Di-8-ANEPPS (non-ratio), DiA (4-Di-16-ASP), Dichlorodihydrofluorescein Diacetate (DCFH), DiD - Lipophilic Tracer, DiD (DiIC18(5)), DIDS, Dihydorhodamine 123 (DHR), DiI (DiIC18(3)), Dinitrophenol, DiO (DiOC18(3)), DiR, DiR (DiIC18(7)), DM-NERF (high pH), DNP, Dopamine, DTAF, DY-630-NHS, DY-635-NHS, DyLight 405, DyLight 488, DyLight 549, DyLight 633, DyLight 649, DyLight 680, DyLight 800, ELF 97, Eosin, Erythrosin, Erythrosin ITC, Ethidium Bromide, Ethidium homodimer -1 (EthD-1), Euchrysin, EukoLight, Europium (III) chloride, Fast Blue, FDA, Feulgen (Pararosaniline), FIF (Formaldehyd Induced Fluorescence), FITC, Flazo Orange, Fluo-3, Fluo-4, Fluorescein (FITC), Fluorescein Diacetate, Fluoro-Emerald, Fluoro-Gold (Hydroxystilbamidine), Fluor-Ruby, FluorX, FM 1-43, FM 4-46, Fura Red (high pH), Fura Red/Fluo-3, Fura-2, Fura-2/BCECF, Genacryl Brilliant Red B, Genacryl Brilliant Yellow 10GF, Genacryl Pink 3G, Genacryl Yellow 5GF, GeneBlazer (CCF2), Gloxalic Acid, Granular blue, Haematoporphyrin, Hoechst 33258, Hoechst 33342, Hoechst 34580, HPTS,

Hydroxycoumarin, Hydroxystilbamidine (FluoroGold), Hydroxytryptamine, Indo-1, high calcium, Indo-1, low calcium, Indodicarbocyanine (DiD), Indotricarbocyanine (DiR), Intrawhite Cf, JC-1, JO-JO-1, JO-PRO-1, LaserPro, Laurodan, LDS 751 (DNA), LDS 751 (RNA), Leucophor PAF, Leucophor SF, Leucophor WS, Lissamine Rhodamine, Lissamine Rhodamine B, Calcein/Ethidium homodimer, LOLO-1, LO-PRO-1, Lucifer Yellow, Lyso Tracker Blue, Lyso Tracker Blue-White, Lyso Tracker Green, Lyso Tracker Red, Lyso Tracker Yellow, LysoSensor Blue, LysoSensor Green, LysoSensor Yellow/Blue, Mag Green, Magdala Red (Phloxin B), Mag-Fura Red, Mag-Fura-2, Mag-Fura-5, Mag-Indo-1, Magnesium Green, Magnesium Orange, Malachite Green, Marina Blue, Maxilon Brilliant Flavin 10 GFF, Maxilon Brilliant Flavin 8 GFF, Merocyanin, Methoxycoumarin, Mitotracker Green FM, Mitotracker Orange, Mitotracker Red, Mitramycin, Monobromobimane, Monobromobimane (mBBr-GSH), Monochlorobimane, MPS (Methyl Green Pyronine Stilbene), NBD, NBD Amine, Nile Red, Nitrobenzoxadidole, Noradrenaline, Nuclear Fast Red, Nuclear Yellow, Nylosan Brilliant Iavin E8G, Oregon Green, Oregon Green 488-X, Oregon Green, Oregon Green 488, Oregon Green 500, Oregon Green 514, Pacific Blue, Pararosaniline (Feulgen), PBFI, Phloxin B (Magdala Red), Phorwite AR, Phorwite BKL, Phorwite Rev, Phorwite RPA, Phosphine 3R, PKH26 (Sigma), PKH67, PMIA, Pontochrome Blue Black, POPO-1, POPO-3, PO-PRO-1, PO-PRO-3, Primuline, Procion Yellow, Propidium Iodid (PI), PyMPO, Pyrene, Pyronine, Pyronine B, Pyrozal Brilliant Flavin 7GF, QSY 7, Quinacrine Mustard, Resorufin, RH 414, Rhod-2, Rhodamine, Rhodamine 110, Rhodamine 123, Rhodamine 5 GLD, Rhodamine 6G, Rhodamine B, Rhodamine B 200, Rhodamine B extra, Rhodamine BB, Rhodamine BG, Rhodamine Green, Rhodamine Phallicidine, Rhodamine Phalloidine, Rhodamine Red, Rhodamine WT, Rose Bengal, S65A, S65C, S65L, S65T, SBFI, Serotonin, Sevron Brilliant Red 2B, Sevron Brilliant Red 4G, Sevron Brilliant Red B, Sevron Orange, Sevron Yellow L, SITS, SITS (Primuline), SITS (Stilbene Isothiosulphonic Acid), SNAFL calcein, SNAFL-1, SNAFL-2, SNARF calcein, SNARF1, Sodium Green, SpectrumAqua, SpectrumGreen, SpectrumOrange, Spectrum Red, SPQ (6-methoxy-N-(3-sulfopropyl)quinolinium), Stilbene, Sulphorhodamine B can C, Sulphorhodamine Extra, SYTO 11, SYTO 12, SYTO 13, SYTO 14, SYTO 15, SYTO 16, SYTO 17, SYTO 18, SYTO 20, SYTO 21, SYTO 22, SYTO 23, SYTO 24, SYTO 25, SYTO 40, SYTO 41, SYTO 42, SYTO 43, SYTO 44, SYTO 45, SYTO 59, SYTO 60, SYTO 61, SYTO 62, SYTO 63, SYTO 64, SYTO 80, SYTO 81, SYTO 82, SYTO 83, SYTO 84, SYTO 85, SYTOX Blue, SYTOX Green, SYTOX Orange, Tetracycline, Tetramethylrhodamine (TAMRA), Texas Red, Texas Red-X conjugate, Thiadicarbocyanine (DiSC3), Thiazine Red

R, Thiazole Orange, Thioflavin 5, Thioflavin S, Thioflavin TCN, Thiolyte, Thiozole Orange, Tinopol CBS (Calcofluor White), TMR, TO-PRO-1, TO-PRO-3, TO-PRO-5, TOTO-1, TOTO-3, TRITC (tetramethylrodamine isothiocyanate), True Blue, TruRed, Ultralite, Uranine B, Uvitex SFC, WW 781, X-Rhodamine, XRITC, Xylene Orange, Y66F, Y66H, Y66W, YO-PRO-1, YO-PRO-3, YOYO-1, YOYO-3, SYBR Green, Thiazole orange (interchelating dyes), or combinations thereof.

Another aspect of the invention is directed to a computer-implemented method. For instance, a computer and/or an automated system may be provided that is able to automatically and/or repetitively perform any of the methods described herein. As used herein, "automated" devices refer to devices that are able to operate without human direction, i.e., an automated device can perform a function during a period of time after any human has finished taking any action to promote the function, e.g. by entering instructions into a computer to start the process. Typically, automated equipment can perform repetitive functions after this point in time. The processing steps may also be recorded onto a machine-readable medium in some cases.

For example, in some cases, a computer may be used to control imaging of the sample, e.g., using fluorescence microscopy, STORM or other super-resolution techniques such as those described herein. In some cases, the computer may also control operations such as drift correction, physical registration, hybridization and cluster alignment in image analysis, cluster decoding (e.g., fluorescent cluster decoding), error detection or correction (e.g., as discussed herein), noise reduction, identification of foreground features from background features (such as noise or debris in images), or the like. As an example, the computer may be used to control activation and/or excitation of signaling entities within the sample, and/or the acquisition of images of the signaling entities. In one set of embodiments, a sample may be excited using light having various wavelengths and/or intensities, and the sequence of the wavelengths of light used to excite the sample may be correlated, using a computer, to the images acquired of the sample containing the signaling entities. For instance, the computer may apply light having various wavelengths and/or intensities to a sample to yield different average numbers of signaling entities in each region of interest (e.g., one activated entity per location, two activated entities per location, etc.). In some cases, this information may be used to construct an image and/or determine the locations of the signaling entities, in some cases at high resolutions, as noted above.

In some aspects, the sample is positioned on a microscope. In some cases, the microscope may contain one or more channels, such as microfluidic channels, to direct or

control fluid to or from the sample. For instance, in one embodiment, nucleic acid probes such as those discussed herein may be introduced and/or removed from the sample by flowing fluid through one or more channels to or from the sample. In some cases, there may also be one or more chambers or reservoirs for holding fluid, e.g., in fluidic communication with the channel, and/or with the sample. Those of ordinary skill in the art will be familiar with channels, including microfluidic channels, for moving fluid to or from a sample.

As used herein, "microfluidic," "microscopic," "microscale," the "micro-" prefix (for example, as in "microchannel"), and the like generally refers to elements or articles having widths or diameters of less than about 1 mm, and less than about 100 microns (micrometers) in some cases. In some embodiments, larger channels may be used instead of, or in conjunction with, microfluidic channels for any of the embodiments discussed herein. For examples, channels having widths or diameters of less than about 10 mm, less than about 9 mm, less than about 8 mm, less than about 7 mm, less than about 6 mm, less than about 5 mm, less than about 4 mm, less than about 3 mm, or less than about 2 mm may be used in certain instances. In some cases, the element or article includes a channel through which a fluid can flow. In all embodiments, specified widths can be a smallest width (i.e. a width as specified where, at that location, the article can have a larger width in a different dimension), or a largest width (i.e. where, at that location, the article has a width that is no wider than as specified, but can have a length that is greater). Thus, for instance, the microfluidic channel may have an average cross-sectional dimension (e.g., perpendicular to the direction of flow of fluid in the microfluidic channel) of less than about 1 mm, less than about 500 microns, less than about 300 microns, or less than about 100 microns. In some cases, the microfluidic channel may have an average diameter of less than about 60 microns, less than about 50 microns, less than about 40 microns, less than about 30 microns, less than about 25 microns, less than about 10 microns, less than about 5 microns, less than about 3 microns, or less than about 1 micron.

A "channel," as used herein, means a feature on or in an article (e.g., a substrate) that at least partially directs the flow of a fluid. In some cases, the channel may be formed, at least in part, by a single component, e.g. an etched substrate or molded unit. The channel can have any cross-sectional shape, for example, circular, oval, triangular, irregular, square or rectangular (having any aspect ratio), or the like, and can be covered or uncovered (i.e., open to the external environment surrounding the channel). In embodiments where the channel is completely covered, at least one portion of the channel can have a cross-section that is

completely enclosed, and/or the entire channel may be completely enclosed along its entire length with the exception of its inlet and outlet.

A channel may have any aspect ratio, e.g., an aspect ratio (length to average cross-sectional dimension) of at least about 2:1, more typically at least about 3:1, at least about 5:1, at least about 10:1, etc. As used herein, a "cross-sectional dimension," in reference to a fluidic or microfluidic channel, is measured in a direction generally perpendicular to fluid flow within the channel. A channel generally will include characteristics that facilitate control over fluid transport, e.g., structural characteristics and/or physical or chemical characteristics (hydrophobicity vs. hydrophilicity) and/or other characteristics that can exert a force (e.g., a containing force) on a fluid. The fluid within the channel may partially or completely fill the channel. In some cases the fluid may be held or confined within the channel or a portion of the channel in some fashion, for example, using surface tension (e.g., such that the fluid is held within the channel within a meniscus, such as a concave or convex meniscus). In an article or substrate, some (or all) of the channels may be of a particular size or less, for example, having a largest dimension perpendicular to fluid flow of less than about 5 mm, less than about 2 mm, less than about 1 mm, less than about 500 microns, less than about 200 microns, less than about 100 microns, less than about 60 microns, less than about 50 microns, less than about 40 microns, less than about 30 microns, less than about 25 microns, less than about 10 microns, less than about 3 microns, less than about 1 micron, less than about 300 nm, less than about 100 nm, less than about 30 nm, or less than about 10 nm or less in some cases. In one embodiment, the channel is a capillary.

A variety of materials and methods, according to certain aspects of the invention, can be used to form devices or components containing microfluidic channels, chambers, etc. For example, various devices or components can be formed from solid materials, in which the channels can be formed via micromachining, film deposition processes such as spin coating and chemical vapor deposition, physical vapor deposition, laser fabrication, photolithographic techniques, etching methods including wet chemical or plasma processes, electrodeposition, and the like. See, for example, *Scientific American*, 248:44-55, 1983 (Angell, *et al*).

In one set of embodiments, various structures or components can be formed of a polymer, for example, an elastomeric polymer such as polydimethylsiloxane ("PDMS"), polytetrafluoroethylene ("PTFE" or Teflon®), or the like. For instance, according to one embodiment, a channel such as a microfluidic channel may be implemented by fabricating the fluidic system separately using PDMS or other soft lithography techniques (details of soft lithography techniques suitable for this embodiment are discussed in the references entitled

"Soft Lithography," by Younan Xia and George M. Whitesides, published in the *Annual Review of Material Science*, 1998, Vol. 28, pages 153-184, and "Soft Lithography in Biology and Biochemistry," by George M. Whitesides, Emanuele Ostuni, Shuichi Takayama, Xingyu Jiang and Donald E. Ingber, published in the *Annual Review of Biomedical Engineering*, 2001, Vol. 3, pages 335-373; each of these references is incorporated herein by reference).

Other examples of potentially suitable polymers include, but are not limited to, polyethylene terephthalate (PET), polyacrylate, polymethacrylate, polycarbonate, polystyrene, polyethylene, polypropylene, polyvinylchloride, cyclic olefin copolymer (COC), polytetrafluoroethylene, a fluorinated polymer, a silicone such as polydimethylsiloxane, polyvinylidene chloride, bis-benzocyclobutene ("BCB"), a polyimide, a fluorinated derivative of a polyimide, or the like. Combinations, copolymers, or blends involving polymers including those described above are also envisioned. The device may also be formed from composite materials, for example, a composite of a polymer and a semiconductor material.

In some embodiments, various microfluidic structures or components of the device are fabricated from polymeric and/or flexible and/or elastomeric materials, and can be conveniently formed of a hardenable fluid, facilitating fabrication via molding (e.g. replica molding, injection molding, cast molding, etc.). The hardenable fluid can be essentially any fluid that can be induced to solidify, or that spontaneously solidifies, into a solid capable of containing and/or transporting fluids contemplated for use in and with the fluidic network. In one embodiment, the hardenable fluid comprises a polymeric liquid or a liquid polymeric precursor (i.e. a "prepolymer"). Suitable polymeric liquids can include, for example, thermoplastic polymers, thermoset polymers, waxes, metals, or mixtures or composites thereof heated above their melting point. As another example, a suitable polymeric liquid may include a solution of one or more polymers in a suitable solvent, which solution forms a solid polymeric material upon removal of the solvent, for example, by evaporation. Such polymeric materials, which can be solidified from, for example, a melt state or by solvent evaporation, are well known to those of ordinary skill in the art. A variety of polymeric materials, many of which are elastomeric, are suitable, and are also suitable for forming molds or mold masters, for embodiments where one or both of the mold masters is composed of an elastomeric material. A non-limiting list of examples of such polymers includes polymers of the general classes of silicone polymers, epoxy polymers, and acrylate polymers. Epoxy polymers are characterized by the presence of a three-membered cyclic ether group commonly referred to as an epoxy group, 1,2-epoxide, or oxirane. For example, diglycidyl

ethers of bisphenol A can be used, in addition to compounds based on aromatic amine, triazine, and cycloaliphatic backbones. Another example includes the well-known Novolac polymers. Non-limiting examples of silicone elastomers suitable for use according to the invention include those formed from precursors including the chlorosilanes such as methylchlorosilanes, ethylchlorosilanes, phenylchlorosilanes, etc.

Silicone polymers are used in certain embodiments, for example, the silicone elastomer polydimethylsiloxane. Non-limiting examples of PDMS polymers include those sold under the trademark Sylgard by Dow Chemical Co., Midland, MI, and particularly Sylgard 182, Sylgard 184, and Sylgard 186. Silicone polymers including PDMS have several beneficial properties simplifying fabrication of various structures of the invention. For instance, such materials are inexpensive, readily available, and can be solidified from a prepolymeric liquid via curing with heat. For example, PDMSs are typically curable by exposure of the prepolymeric liquid to temperatures of about, for example, about 65 °C to about 75 °C for exposure times of, for example, at least about an hour. Also, silicone polymers, such as PDMS, can be elastomeric and thus may be useful for forming very small features with relatively high aspect ratios, necessary in certain embodiments of the invention. Flexible (e.g., elastomeric) molds or masters can be advantageous in this regard.

One advantage of forming structures such as microfluidic structures or channels from silicone polymers, such as PDMS, is the ability of such polymers to be oxidized, for example by exposure to an oxygen-containing plasma such as an air plasma, so that the oxidized structures contain, at their surface, chemical groups capable of cross-linking to other oxidized silicone polymer surfaces or to the oxidized surfaces of a variety of other polymeric and non-polymeric materials. Thus, structures can be fabricated and then oxidized and essentially irreversibly sealed to other silicone polymer surfaces, or to the surfaces of other substrates reactive with the oxidized silicone polymer surfaces, without the need for separate adhesives or other sealing means. In most cases, sealing can be completed simply by contacting an oxidized silicone surface to another surface without the need to apply auxiliary pressure to form the seal. That is, the pre-oxidized silicone surface acts as a contact adhesive against suitable mating surfaces. Specifically, in addition to being irreversibly sealable to itself, oxidized silicone such as oxidized PDMS can also be sealed irreversibly to a range of oxidized materials other than itself including, for example, glass, silicon, silicon oxide, quartz, silicon nitride, polyethylene, polystyrene, glassy carbon, and epoxy polymers, which have been oxidized in a similar fashion to the PDMS surface (for example, via exposure to an oxygen-containing plasma). Oxidation and sealing methods useful in the context of the

present invention, as well as overall molding techniques, are described in the art, for example, in an article entitled "Rapid Prototyping of Microfluidic Systems and Polydimethylsiloxane," *Anal. Chem.*, 70:474-480, 1998 (Duffy *et al.*), incorporated herein by reference.

The following documents are each incorporated herein by reference in their entireties: U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*; and Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*

In addition, incorporated herein by reference in their entireties are U.S. Provisional Patent Application Serial No. 62/031,062, filed July 30, 2014, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/050,636, filed September 15, 2014, entitled "Probe Library Construction," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/142,653, filed April 3, 2015, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; and a PCT application filed on even date herewith, entitled "Probe Library Construction," by Zhuang, *et al.*

The following examples are intended to illustrate certain embodiments of the present invention, but do not exemplify the full scope of the invention.

EXAMPLE 1

The example presents a platform to enable the simultaneous detection of the number and spatial organization of thousands of distinct mRNAs within single cells with high efficiency and low error-rate using a novel form of highly multiplexed fluorescence *in situ* hybridization (FISH). This example accomplishes these measurements by integrating and innovating methods for massively parallel probe synthesis, super-resolution imaging, and self-correcting error-checking codes.

Here, these examples present methods for the simultaneous detection of some or all of the thousands of unique RNAs expressed in a cell. This approach not only promises to revolutionize the throughput of the already effective single-molecule FISH (smFISH) approach, but also allows researchers to benefit from the hypothesis free discovery approach which has made other whole-genome systems approaches to biology so effective. For example, this whole genome approach may allow researchers to discover RNAs whose expression levels and/or subcellular localization patterns change under certain conditions of interest, such as disease states, without knowing, *a priori*, which mRNA will change in

abundance or localization. Simultaneous measurements of hundreds of genes within a single cell also allow for the identification of correlations between genes in expression and localization patterns in some cases.

This can be achieved using methods for highly multiplexed smFISH via the sequential hybridization of orthogonal detection probes and super-resolution imaging, reducing the cost of probe synthesis, and the development of a highly automated system to minimize demands on the user, as discussed herein. This provides an integrated platform to handle the bioinformatics of probe design, the mathematics of error-correcting codes, the complexity of image registration and analysis, and the cumbersome fluid handling through a simple suite of user friendly interfaces. This integration allows easy operation with limited user training and facilitates the rapid collection of data.

This example illustrates: (1) computational design of "codewords" to attach to all RNA targets in the cell that will allow unique identification of each RNA with some degree of experimental error tolerance, (2) translation of these codewords into nucleotide sequences and synthesis of the required single-stranded (ss) oligonucleotide (e.g. ssDNA) probes, (3) sample fixation and hybridization of these probes to the RNA targets *in situ*, (4) read-out of these codewords via successive rounds of hybridization of distinct fluorescent probes imaged with conventional fluorescence microscopy or super-resolution fluorescence microscopy, and (5) automated decoding of measured codewords combined with computational error correction to uniquely and robustly identify individual mRNAs.

In the first step, a "codeword" is assigned to every RNA that is to be labeled. In a typical design these may be strings of N binary letters or positions. Codewords may be chosen from the same wide range of existing error tolerant or error-correcting encoding schemes developed for digital storage and communication, e.g., using Hamming codes or the like. For example, actin-RNA may be assigned the binary codeword 11001010. Each codeword may be unique and separated from the other codewords by a Hamming distance, h, which measures the number of letters or positions that must be incorrectly read for one codeword to be misinterpreted as a different one. A Hamming distance greater than 1 between all codewords allows for some measurement errors to be detected—since simple errors would produce codewords that are not used to encode RNAs. For a Hamming distance larger than 2, it is also possible to correct some errors, as codewords with one error will be closest in Hamming distance to a single, unique codeword. The total number of different RNAs to be detected from the transcriptome and the amount of error correction desired

determines the length of the codewords. Information theory provides several efficient algorithms for assembling error-correcting binary codebooks.

In the second step, this encoding scheme is translated into a set of oligonucleotide (e.g. DNA) probe sequences, which can be called primary probes or encoding probes, each of which not only targets a probe to the RNA of interest but also encodes the unique binary codeword within a set of secondary binding sites (Fig. 1). For example, first designed may be primary binding sequences for each targeted mRNA. These sequences are "target sequences" that are comprised of complementary nucleotide sequences to their target RNAs computationally selected to satisfy a stringent set of hybridization conditions, including uniqueness in the target genome. To improve the efficiency of hybridization to individual mRNAs, multiple primary target sequences are designed for each individual RNA. Then each position within the set of codewords is assigned a unique oligonucleotide (e.g. DNA) sequence, which is called a read sequence. These tags are designed as to have no interaction with endogenous mRNA sequences or each other. For instance, for all the value "1"s in a codeword of an individual mRNA, the corresponding read sequence is attached to the primary targeting sequences against that mRNA. In general, each probe will contain a target sequence and one or more read sequences. If the total length of the necessary read sequences and the primary target sequence exceeds synthesis capabilities, then subsets of the read sequences can be appended to distinct target sequences. For example, consider the potential codeword 11001010 for actin. Probe sequences for this RNA could contain the read sequences corresponding to positions 1, 2, 5, and 7 in the codeword attached to a variety of target sequences specific to actin. After all the sequences have been designed, the resulting complex set of unique custom oligonucleotide (e.g. DNA) sequences is manufactured and amplified using methods as described below.

In the third step, the resulting pool of DNA is hybridized, e.g., to fixed, permeabilized cells. In this process, individual probes may be attached to every RNA in the cell by hybridization of their corresponding target sequences with the RNA while the read sequences remain free to bind the appropriate secondary probes as discussed below.

In the fourth step—the read-out step—fluorescently labeled secondary nucleic acid probes (also called readout probes) are successively hybridized to the read sequences attached to the target sequences that binds to the mRNA targets in the above step. When simultaneously imaging a large number of different RNA species in cells, the density of labeled RNAs may exceed that at which each RNA can be resolved via conventional imaging methods. Thus, this may be performed using a super-resolution imaging method, for example

STORM (stochastic optical reconstruction microscopy), to resolve the labeled molecules. After each round of hybridization and imaging with the secondary probes, the fluorophores are quenched or otherwise inactivated either via chemical or optical techniques such as oxidation, chemically bleaching, photobleaching, stringent washing or enzymatic digestion, etc. The sample is then stained with the next secondary probe, and the cycle continues until all positions of the codewords have been read out. In the simplest incarnation, there will be one hybridization step for each position within the codeword, e.g. 8 hybridization steps for an 8-letter codeword (Fig. 1).

Fig. 1 shows schematic diagrams of this example. Fig. 1A shows that every position of the codewords is assigned a unique oligonucleotide sequence when this position has a value "1." All mRNA codewords are then translated into combinations of read sequences, which are attached to the targeting sequence. Fig. 1B shows various steps of the labeling scheme of this example. In the first step, all mRNAs (I - III) are tagged with multiple oligonucleotide (e.g. ssDNA) probes comprising a primary targeting sequence, which hybridizes to the RNA of interest, and a "tail" (i.e. containing read sequences) carrying the translated codeword, which does not interact with endogenous nucleotide sequences. In the next step, the first secondary probe is added, which can bind all probes whose tails have a read sequence corresponding to the value of "1" in the first position. The dyes on these secondary probes are imaged and bleached, then the next secondary probe is added to bind probes attached to mRNA which have a value of "1" in the second position of their assigned codeword, and so on.

In the final step, the microscopy images from each staining and imaging round are aligned, for example, computationally (e.g. using fiducial beads or other markers tracked during image acquisition), and the clusters of localizations resolved by conventional fluorescence microscopy or super-resolution imaging (e.g. STORM) from the different rounds are identified. These clusters of localizations arise from individual target mRNA molecules, and the hybridization rounds in which a spot was detected in a given cluster correspond to the "1" in the codeword for that mRNA. If there are no missed-detection events or false positive signal in the images, this codeword will perfectly match one of the expected codewords. Fig. 1 describes an example in which the codeword has three letters, i.e. three positions, and the three target mRNAs have codewords 110, 101, and 011 assigned to them. In real experimental examples, the codeword could contain more digits. For example, the mRNA for actin can be assigned the codeword, 11001010. In that case, detected clusters containing overlapping localization signals in the 1st, 2nd, 5th and 7th hybridization steps

(meaning the 1st, 2nd, 5th and 7th secondary probes bound to this site) can be identified as individual actin mRNA molecules, since the pattern of positive bindings matches the codewode of actin (11001010). In addition, if there are missed-detection events or false positive signals in the image data, these aberrations can be corrected by the implemented error-correction scheme. For example, clusters of localizations with a detected codeword that has only one digit discrepancy from 11001010 (such as 11000010 or 11101010) can also be identified as actin mRNA since all other valid codeword in this example differ from the detected pattern in two more positions.

EXAMPLE 2

This example describes another alternative approach that differs in several of the steps described above. This approach begins with the first step, construction of the codewords to the desired mRNA targets, as described above.

In the second step of this approach, nucleic acid probes are designed that bind uniquely to the mRNA targets of interest, as described above. However, instead of appending unique read sequences to these targeting sequences, unique pools or groups of probes are constructed from these target sequences. Each pool comprises all or a subset of the sequences that target all mRNAs which contain the same value at a given position in their codeword. For example, the first pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the first position of their codewords, e.g. 110 and 101 but not 011; the second pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the second position of their codewords, e.g. 110 and 011 but not 101; the third pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the third position of their codewords, e.g. 011 and 101 but not 110 (Fig. 1C). As another example, consider the potential codeword 11001010 for actin. Probes that target this mRNA would be included in pools 1, 2, 5, and 7 but not in pools 3, 4, 6, and 8. The same target to a given mRNA may or may not be included in pools. For example, a probe that targets the same region of actin may be included in pools 1, 2, 5, and 7 or any subset of these pools. After all of the pools have been designed, each complex set of unique, custom oligonucleotide sequences is manufactured and amplified using methods as described below.

In the third step of this approach, the first pool of probes is hybridized, e.g., to fixed, permeabilized cells. In this process, the fluorophores attached to each of the probes in this pool are bound to each of the targets of that pool. The binding of these probes is then determined by fluorescence microscopy. As described above, these images can be collected either via a range of methods including both conventional fluorescence imaging or super-

resolution imaging methods such as STORM. After a round of imaging, the probes from the first pool are inactivated or removed from the sample via the methods described above. This process is then repeated for each successive pool of probes until some or all of the pools have been applied to the sample and imaged such that all positions in the codewords have been read out. In the simplest incarnation, there will be one hybridization and imaging step for each position in the codeword, e.g. 3 rounds of hybridization and imaging for a codeword with 3 positions (Fig 1C) or 8 rounds of hybridization and imaging for a codeword with 8 positions.

The final step of this approach is identical to that described above.

EXAMPLE 3

In this example, 14 genes (PGK1, H3F3B, PKM, ENO1, GPI, EEF2, GNAS, HSPA8, GAPDH, CALM1, RHOA, PPIA, UBA52, and VCP) were encoded using a subset of the (8,4) SECDED code (Figs. 2A-2E). To determine the accuracy of these measurements, the measured abundances of these 14 mRNAs were compared to the abundances measured from bulk RNA-seq of A549 cells (published ENCODE data). Remarkably, it was found that there was excellent agreement between these two measurements, as the transcripts count measured using the sequential hybridization approach correlated with gene expression measured using RNA-seq with a Pearson correlation coefficient r of 0.75 (Fig. 2F). Gene expression from 3 other cells were also measured, and it was found that the gene expression of these 14 genes was highly correlated among the cells with a r of 0.96 (Fig. 2G).

Codebook Design. Each mRNA in the target set was assigned a binary codeword using a Single Error Correction Double Error Detection (SECDED) code. SECDED is an extended Hamming codebook with an additional parity bit. Briefly, Matlab's Communications System toolbox was used to generate SECDED codes of either 8 or 16 letters or positions. In both cases, only those codewords containing four 1s were used. These words were assigned at random to mRNAs in the target set. [0 1 0 1 1 1 0 0] is an example of the 8-letter codewords used (i.e., these codewords each contained four 1s and four 0s.) [0 1 0 1 1 1 0 0 0 0 0 0 0 0 0 0] is an example of the 16-letter codewords used (i.e., each codeword contained four 1s and twelve 0s). Not every codeword was necessarily assigned to an mRNA.

Computational Assembly of ssDNA Primary Probe Sequence. The number of primary nucleic acid probes used for hybridization with mRNA targets ranged from 200 to 2000 unique oligonucelotides, depending on the experiment. For example, to label 14 mRNAs with 28 oligos targeting each gene, 392 unique sequences were used. Large number

of oligos with unique sequences were purchased in a pool from LC Sciences or CustomArray. However, array synthesized oligos were in minute quantity that was insufficient for *in situ* hybridization. The protocol for their amplification are described below.

Each primary probe contained three components: flanking primer sequences to allow enzymatic amplification of probes, targeting sequence for *in situ* hybridization to mRNAs, and secondary tag sequence containing one or more read sequences for sequential readout of codewords.

The following is an example of a primary probe:

GTTGGCGACGAAAGCACTGCGATTGGAACCGTCCCAAGCGTTGCG
CTTAATGGATCATCAATTTTGTCTCACTACGACGGTCAATCGCGCTGCATACTTG
CGTCGGTCGGACAAACGAGG

(SEQ ID NO: 1)

The components are arranged in the following order: forward primer (not underlined), secondary read sequence 1 (underlined), mRNA targeting sequence (not underlined), secondary read sequence 2 (underlined), and reverse primer (not underlined). The secondary read sequences are the reverse complement of the corresponding secondary probes. Since only codewords that contained four '1's were used, the primary probes for each mRNA needed to contain 4 different secondary read sequences in this example. However, in order to reduce the overall length of the primary probes, the pool of targeting sequences for each mRNA target was split at random into two pools. Two secondary read sequences are attached each probe in one of the two pools and the other two secondary read sequences are attached the probes in the other pool. The design criteria for each component are described below.

Primer Design. Specific index primers were generated by a collection of 240,000 published sequences of orthogonal 25-bp long sequences. These sequences were trimmed to 20 bp, selected for a narrow 70 to 80 °C melting temperature, the absence of consecutive repeats of 3 or more base, and the presence of a GC clamp, i.e. one of the two 3' terminal bases must be G or C. To further improve specificity, these sequences were then screened against the human genome using BLAST+ (Camacho et al 2009), and primers with 14 or more contiguous bases of homology were eliminated. In a subsequent screening via BLAST+, primers that shared 11 or more contiguous bases or more than 5 bases at the 3' end of any other primer or the T7 promoter were also removed.

Secondary Probes Design. 30-bp long secondary probe sequences were created by concatenating fragments of the orthogonal primer set described above. These secondaries were then screened for orthogonality with other secondaries (no more than 11 basepairs of homology) and for potential off-target binding sites in the human genome (no more than 14 basepairs of homology). Secondary sequences used in this example are provided in Table 1.

Table 1

Bit	Secondary sequences	Sequence number
B1	CGCAACGCTTGGGACGGTTCCAATCGGATC	SEQ ID NO: 2
B2	CGAATGCTCTGGCCTCGAACGAACGATAGC	SEQ ID NO: 3
B3	ACAAATCCGACCAGATCGGACGATCATGGG	SEQ ID NO: 4
B4	CAAGTATGCAGCGCGATTGACCGTCTCGTT	SEQ ID NO: 5
B5	TGCGTCGTCTGGCTAGCACGGCACGCAAAT	SEQ ID NO: 6
B6	AAGTCGTACGCCGATGCGCAGCAATTCACT	SEQ ID NO: 7
B7	CGAAACATCGGCCACGGTCCCGTTGAACTT	SEQ ID NO: 8
B8	ACGAATCCACCGTCCAGCGCGTCAAACAGA	SEQ ID NO: 9
B9	CGCGAAATCCCCGTAACGAGCGTCCCTTGC	SEQ ID NO: 10
B10	GCATGAGTTGCCTGGCGTTGCGACGACTAA	SEQ ID NO: 11
B11	CCGTCGTCTCCGGTCCACCGTTGCGCTTAC	SEQ ID NO: 12
B12	GGCCAATGGCCCAGGTCCGTCACGCAATTT	SEQ ID NO: 13
B13	TTGATCGAATCGGAGCGTAGCGGAATCTGC	SEQ ID NO: 14
B14	CGCGCGGATCCGCTTGTCGGGAACGGATAC	SEQ ID NO: 15
B15	GCCTCGATTACGACGGATGTAATTCGGCCG	SEQ ID NO: 16
B16	GCCCGTATTCCCGCTTGCGAGTAGGGCAAT	SEQ ID NO: 17

mRNA Targeting Sequence Design. To determine the relative abundance of all the isoforms of all genes expressed in these cell lines, transcriptome profiling data from the ENCODE project for total RNA from A549 and IMR90 cells was processed using the publicly available software *cufflinks*, along with human genome annotations from gencode v18. Gene models corresponding to the most highly expressed isoform were used to build a sequence library in FASTA format recording the dominant isoform of every gene. Genes of interest were selected from this library. These genes were partitioned into 1kb segments, then

the software *OligoArray2.1* was used to generate primary probe sequences for the human transcriptome with the following constraints: 30-bp or 40-bp length, depending on the experiment; probe-target melting temperatures greater than 70 °C (variable parameter); no cross hybridization targets with melting temperatures greater than 72 °C (variable parameter); no predicted internal secondary structures with melting temperatures greater than 76 °C (variable parameter); and no single-nucleotide contiguous repeats of 6 or more bases. After OligoArray probe selection, all potential probes that mapped to a different gene were rejected while all potential probes with multiple alignments to the same gene were retained. A BLAST database was assembled from the FASTA library of all expressed genes to screen for probes' uniqueness. For each gene, 14 to 28 targeting sequences produced during the OligoArray processing were selected.

Probe Synthesis – index PCR. The template for specific probe sets were selected from the complex oligopool via limited-cycle PCR. Briefly, 0.5 to1 ng of the complex oligopool was combined with 0.5 micromolar of each primer. The forward primer matched the priming sequence for the desired subset while the reverse primer was a 5' concatenation of this sequence with a T7 promoter. To avoid the generation of G-quadruplets, which can be difficult to synthesize, the terminal Gs required in the T7 promoter were generated from Gs located at the 5' of the priming region where appropriate. All primers were synthesized by IDT. A 50 microliter reaction volume was amplified either using the KAPA real-time library amplification kit (KAPA Biosystems; KK2701) or via a homemade qPCR mix which included 0.8X EvaGreen (Biotum; 31000-T) and the hot-start Phusion polymerase (New England Biolabs; M0535S). Amplification was followed in real time using Agilent's MX300P or Biorad's CFX Connect. Individual samples were removed immediately before the plateau in amplification to minimize distortion of template abundance due to over-amplification. Individual templates were purified with columns according to the manufacturer's instructions (Zymo DNA Clean and Concentrator; D4003) and eluted in RNase-free deionized water.

Amplification via *in-vitro* transcription. The template was then amplified via *in vitro* transcription. Briefly, 0.5 to 1 microgram of template DNA was amplified into 100-200 microgram of RNA in a single 20-30 microliter reaction with a high yield RNA polymerase (New England Biolabs; E2040S). Reactions were supplemented with 1X RNase inhibitor (Promega RNasin; N2611). Amplification was typically run for 4 to 16 hours at 37 °C to maximize the yield. The RNA was not purified after the reaction and was either stored at -80 °C or immediately converted into DNA as described below.

Reverse Transcription. 1 to 2 nmol of fluorescently-labeled ssDNA probe was created from the above *in vitro* transcription reactions using the reverse transcriptase Maxima H- (Thermo Scientific; EP0751). This enzyme was used because of its higher processivity and temperature resistance, which allowed the conversion of large quantities of RNA into DNA within small volumes at temperatures that disfavor secondary structure formation. The unpurified RNA created above was supplemented with 1.6 mM of each dNTP, 1-2 nmol of fluorescently labeled forward primer, 300 units of Maxima H-, 60 units of RNasin, and a final 1X concentration of the Maxima RT buffer. The final 75 microliter volume was incubated at 50 °C for 60 minutes.

Strand Selection and Purification. The template RNA in the reaction above was then removed from the DNA via alkaline hydrolysis. 75 microliters of 0.25 M EDTA and 0.5 N NaOH were added to each reverse transcription reaction, and the sample was incubated at 95 °C for 10 minutes. The reaction was immediately neutralized by purifying the ssDNA probe with a modified version of the Zymo Oligo Clean and Concentrator protocol. Specifically, the 5-microgram capacity column was replaced with a 25-microgram or 100-microgram capacity DNA column as appropriate. The remainder of the protocol was run according to the manufacturer's instructions. Probe was eluted in 100 microliter RNase-free deionized water and evaporated in a vacuum concentrator. The final pellet was resuspended in 10 microliters of RNase-free water and stored at -20 °C. Denaturing poly-acrylimide gel electrophoresis and absorption spectroscopy revealed that this protocol typically produced 90-100% incorporation of the fluorescent primer into full length probe and 75-90% recovery of the total fluorescent probe. Thus, without exceeding a 150-microliter reaction volume, this protocol can be used to create ~2 nmol of fluorescent probe.

Cell culture and fixation. A549 and IMR90 cells (American Type Culture Collection) were cultured with Dulbecco's Modified Eagle Medium and Eagle's Minimum Essential Medium respectively. Cells were incubated at 37 °C with 5% CO_2 for 36-48 hours. Cells were fixed in 3% paraformaldehyde (Electron Microscopy Sciences) in PBS for 15 minutes, washed with PBS, and permeabilized in 70% ethanol overnight at 4 °C.

Fluorescence In Situ Hybridization (FISH) – primary (encoding) probes. Cells were hydrated in wash buffers (2xSSC, 50% formamide) for 10 minutes, labelled with primary oligos (0.5 nM per sequence) in hybridization buffers (2xSSC, 50% formamide, 1 mg/mL yeast tRNA, and 10% dextran sulfate) overnight at 37 °C, washed with wash buffers at 47 °C for 10 minutes twice, and washed with 2xSSC twice. Fluorescent fiducial beads (Molecular Probes, F-8809) were added at a 1:10,000 dilution in 2xSSC before imaging.

Secondary probes. Secondary (readout) probes (10 nM) were hybridized in secondary hybridization buffers (2xSSC, 20% formamide, and 10% dextran sulfate) to their primary targets for 30 minutes at 37 °C. Cells remained on the microscope stage during the hybridization. An objective heater was used to maintain the temperature at 37 °C. Cells were washed with secondary wash buffers (2xSSC, 20% formamide).

Fluidics and STORM Imaging. Multiple rounds of sequential labeling, washing, imaging, and bleaching were performed on an automated platform consisted of a fluidics setup and a STORM (stochastic optical reconstruction microscopy) microscope. The fluidics setup included a flow chamber (Bioptech FCS2), a peristaltic pump (Rainin Dynamax RP-1), and three computer-controlled 8-way valves (Hamilton MVP and Hamilton HVXM 8-5). This system allowed the automated integration of STORM movie collection and secondary hybridization.

The imaging buffer included, 50 mM Tris (pH 8) 10% (w/v) glucose, 1% βME (2-mercaptoethanol) or 25 mM MEA, with or without 2 mM 1,5-cyclooctadiene, and an oxygen scavenging system (0.5 mg/ml glucose oxidase (Sigma-Aldrich) and 40 microgram/ml catalase (Sigma-Aldrich)). A layer of mineral oil was used to seal the imaging buffer, preventing its acidification over the course of multiple hybridizations.

The STORM setup included an Olympus IX-71 inverted microscope configured for oblique incidence excitation. The samples were continuously illuminated with a 642-nm diode-pumped solid-state laser (VFL-P500-642; MPB communications). A 405-nm solid-state laser (Cube 405-100C; Coherent) was used for activation of dyes. Fluorescence was collected using an Olympus (UPlanSApo 100x, 1.4 NA) objective lens and passed through a custom dichroic, as well as a quad-view beam splitter. All movies were recorded using an EMCCD camera (Andor iAxon 897), imaging at 60Hz. The 512x256 field of view of the camera was split into separate 256x256 pixel movies prior to saving. The left half of this field of view contained the STORM data and the right half contained images of the fluorescent feducial beads. These latter movies were downsampled to 1 Hz prior to saving. During data acquisition, a home-built focus lock was used to maintain a constant focal plane. STORM movies included 20,000 to 30,000 frames in STORM buffer while the bleach movies included 10,000 frames in wash buffer.

Image Analysis – analysis of single-molecule localizations. Movies of single molecule localizations and fluorescent feducial beads were processed separately using a previously published single-emitter localization software.

Image Registration. The starting position of the beads from each round of hybridization were used to align movies from each round. The 2D autocorrelation between bead images of consecutive hybridizations followed by nearest-neighbor matching was used to match beads between images. The pair of beads with the most similar displacement vector were used to compute a rigid translation-rotation warp to align the beads. This alignment method is robust to samples in which multiple feducials are displaced or come detached and reattach during imaging.

Drift Correction. Drift during image acquisition was corrected using the trajectory of the feducial beads (recorded at 1 Hz). Bead positions were linked in each frame. The trajectory of the two beads that moved in the most correlated fashion was taken as the drift trajectory.

mRNA Cluster Calling. Localizations were first screened to be above a threshold number of photons (generally 2000) and required to be within 32 nm of 5 other localizations (parameters may be tuned). The remaining molecule localizations were binned in a 2D histogram of 10x10 nm bins (bin size is a variable parameter). All connected bins were taken to be part of a cluster (diagonal contacts are classified as connected). Clusters were required to have more than 80 total localizations across all hybridizations (variable parameter) to be called an mRNA cluster. The weighted centroids of these clusters from the 2D histogram were recorded as the mRNA positions.

A given cluster is recorded to be represented in an individual hybridization round if more than 9 localizations (variable parameter) are found within a 48 nm radius (variable parameter) of the centroid for that mRNA in each hybridization round.

Cluster Decoding. For each mRNA cluster, a codeword is readout, including "0"s for all the hybridization rounds in which less than the threshold number of localizations are found near the centroid and "1" s for the rounds where above threshold number of localizations are counted. The SECDED codebook decoded these as either perfect matches to target mRNA codewords, correctable errors which can be unambiguously mapped back to target mRNA, or uncorrectable errors, which differed by two or more letters from the words in the codebook.

Fig. 2A shows a STORM image of a cell. Fig. 2B shows a zoom in of the boxed region in Fig. 2A. Each dot indicates a localization. Localizations from different rounds of imaging are shown differently. Fig. 2C shows a representative cluster of localizations from the boxed region in Fig. 2B. The cluster shows localization signals from 4 different hybridizations. This cluster is a putative mRNA encoded with codeword [0 1 0 1 1 1 0 0].

Fig. 2D shows a reconstructed cell image of 14 genes after decoding and error correction. Each gene is shown differently. Fig. 2E shows measured gene expression for the 14 genes from the cell. Fig. 2F shows a comparison of transcript count with ensemble RNA-sequencing data. Fig. 2G shows correlation of transcript expression level between two cells detected using the described approach.

EXAMPLE 4

The following examples are generally directed to multiplexed single-molecule imaging with error-robust encoding allowing for simultaneous measurements of thousands of RNA species in single cells. In general, knowledge of the expression profile and spatial landscape of RNAs in individual cells is essential for understanding the rich repertoire of cellular behaviors. The following examples reports various techniques directed to single-molecule imaging approaches that allow the copy numbers and spatial localizations of thousands of RNA species to be determined in single cells. Some of these techniques are called Multiplexed Error-Robust Fluorescence in Situ Hybridization or "MERFISH."

Using error-robust encoding schemes to combat single-molecule labeling and detection errors, these examples demonstrated the imaging of hundreds to thousands of unique RNA species in hundreds of individual cells. Correlation analysis of the ~10^4 to ~10^6 pairs of genes allowed constraints on gene regulatory networks, prediction of novel functions for many unannotated genes, and identification of distinct spatial distribution patterns of RNAs that correlate with properties of the encoded proteins.

System-wide analyses of the abundance and spatial organization of RNAs in single cells promise to transform understanding in many areas of cell and developmental biology, such as the mechanisms of gene regulation, the heterogeneous behavior of cells, and the development and maintenance of cell fate. Single-molecule fluorescence *in situ* hybridization (smFISH) has emerged as a powerful tool for studying the copy number and spatial organization of RNAs in single cells either in isolation or in their native tissue context. Taking advantage of its ability to map the spatial distributions of specific RNAs with high resolution, smFISH has revealed the importance of subcellular RNA localization in diverse processes such as cell migration, development, and polarization. In parallel, the ability of smFISH to precisely measure the copy numbers of specific RNAs without amplification bias has allowed quantitative measurement of the natural fluctuations in gene expression, which has in turn elucidated the regulatory mechanisms that shape such fluctuations and their role in a variety of biological processes.

However, application of the smFISH approach to many systems-level questions remains limited by the number of RNA species that can be simultaneously measured in single cells. State-of-the-art efforts using combinatorial labeling by either color-based barcodes or sequential hybridization have enabled simultaneous measurements of 10-30 different RNA species in individual cells, yet many interesting biological questions would benefit from the measurement of hundreds to thousands of RNAs within a single cell, which are not achievable using such techniques. For example, analysis of how the expression profile of such a large number of RNAs vary from cell to cell and how these variations correlate among different genes could be used to systematically identify co-regulated genes and map regulatory networks; knowledge of the subcellular organizations of numerous RNAs and their correlations could help elucidate molecular mechanisms underlying the establishment and maintenance of many local cellular structures; and RNA profiling of individual cells in native tissues could allow *in situ* identification of cell type.

The following examples generally discuss certain techniques called MERFISH, which are highly multiplexed smFISH imaging methods that substantially increase the number of RNA species that can be simultaneously imaged in single cells by using combinatorial labeling and sequential imaging with error-robust encoding schemes. These examples demonstrate this multiplexed imaging approach by simultaneously measuring 140 RNA species using an encoding scheme that can both detect and correct errors and 1001 RNA species using an encoding scheme that can detect but not correct errors. It should be understood that these numbers are by way of exemplification only, not limitation. Correlation analyses of the copy number variations and spatial distributions of these genes allowed us to identify groups of genes that are co-regulated and groups of genes that share similar spatial distribution patterns inside the cell.

Combinatorial labeling with error-robust encoding schemes. Combinatorial labeling that identifies each RNA species by multiple (N) distinct signals offers a route to rapidly increase the number of RNA species that can be probed simultaneously in individual cells (Fig. 5A). However, this approach to scaling up the throughput of smFISH to the systems scale faces a significant challenge because not only does the number of addressable RNA species increases exponentially with N, but the detection error rates also grow exponentially with N (Figs. 5B-5D). Imagine a conceptually simple scheme to implement combinatorial labeling, where each RNA species is encoded with a N-bit binary word and the sample is probed with N corresponding rounds of hybridization, each round targeting only the subset of RNAs that should read '1' in the corresponding bit (Fig. 11). N rounds of hybridization

would allow 2^N-1 RNA species to be probed. With just 16 hybridizations, over 64,000 RNA species, which should cover the entire human transcriptome including both messenger RNAs (mRNAs) and non-coding RNAs, could be identified (Fig. 5B; upper symbols). However, as *N* increases, the fraction of RNAs properly detected (the calling rate) would rapidly decrease and, more troublingly, the fraction of RNAs that are identified as incorrect species (the misidentification rate) would rapidly increase (Fig. 5C, lower symbols; Fig. 5D, upper symbols). With realistic error rates per hybridization (measured below), the majority of RNA molecules would be misidentified after 16 rounds of hybridizations!

To address this challenge, error-robust encoding schemes were designed, in which only a subset of the 2^N-1 words separated by a certain Hamming distance were used to encode RNAs. In a codebook where the minimum Hamming distance is 4 (HD4 code), at least four bits must be read incorrectly to change one code word into another (Fig. 12A). As a result, every single-bit error produces a word that is uniquely close to a single code word, allowing such errors to be detected and corrected (Fig. 12B). Double-bit errors produce words with an equal Hamming distance of 2 from multiple code words and, thus, can be detected but not corrected (Fig. 12C). Such a code should substantially increase the calling rate and reduce the misidentification rate (Fig. 5C and 5D, middle symbols). To further account for the fact that it is more likely to miss a hybridization event (an 1-->0 error) than to misidentify a background spot as an RNA (an 0-->1 error) in smFISH measurements, a modified HD4 (MHD4) code was designed, in which the number of '1' bits were kept both constant and relatively low, only four per word, to reduce error and avoid biased detection. This MHD4 code should further increase the calling rate and reduce the misidentification rate (Fig. 5C, upper symbols; Fig. 5D, lower symbols).

In addition to the error considerations, several practical challenges have also made it difficult to probe a large number of RNA species, such as the high cost of the massive number of fluorescently labeled FISH probes needed and the long time required to complete many rounds of hybridization. To overcome these challenges, in this example, a two-step labeling scheme was designed to encode and readout cellular RNAs (Fig. 5E). First, cellular RNAs were labeled with a set of encoding probes (also called primary probes), each probe comprising a RNA targeting sequence and two flanking readout sequences. Four of the *N* unique readout sequences were assigned to each RNA species based on the MHD4 code word of the RNA. Second, these *N* readout sequences were identified with complementary FISH probes, the readout probes (also called secondary probes) via *N* rounds of hybridization and imaging, each round using a unique readout probe. To increase the signal to background

ratio, every cellular RNA was labeled with ~192 encoding probes. Because each encoding probe contained two of the four readout sequences associated with that RNA (Fig. 5E), a maximum of ~96 readout probes can bind to each cellular RNA per hybridization round. To generate the massive number of encoding probes required, they were amplified from array-derived oligonucleotide pools containing tens of thousands of custom sequences using an enzymatic amplification process comprising *in vitro* transcription followed by reverse transcription (Fig. 13, see below regarding probe synthesis). This two-step labeling approach significantly diminished the total hybridization time for an experiment: it was found that efficient hybridization to the readout sequences took only 15 minutes whereas efficient direct hybridization to cellular RNA required more than 10 hours.

Fig. 5 describes MERFISH, a highly multiplexed smFISH approach using combinatorial labeling and error-robust encoding. Fig. 5A shows a schematic depiction of the identification of multiple RNA species in N rounds of imaging. Each RNA species is encoded with a N-bit binary word and during each round of imaging, only the subset of RNAs that should read '1' in the corresponding bit emit signal. Figs. 5B-5D show the number of addressable RNA species (Fig. 5B), the rate at which these RNAs are properly identified (calling rate) (Fig. 5C), and the rate at which RNAs are incorrectly identified as a different RNA species (misidentification rate) (Fig. 5D) plotted as a function of the number of bits (N) in the binary words encoding RNA. In Figs. 5B and 5D, the upper dots are a simple binary code that includes all 2^N-1 possible binary words; the middle dots are the HD4 code where the Hamming distance separating words is 4; and the lower dots are the modified HD4 (MHD4) code where the number of '1' bits are kept at four. These are reversed in Fig. 5C.

The calling and misidentification rates are calculated with per bit error rates of 10% for the 1-->0 error and 4% for the 0-->1 error. Fig. 5E is as schematic diagram of the implementation of a MHD4 code for RNA identification. Each RNA species is first labeled with ~192 encoding probes that convert the RNA into a unique combination of readout sequences (Encoding hyb). These encoding probes each contain a central RNA targeting region flanked by two readout sequences, drawn from a pool of N different sequences, each associated with a specific hybridization round. Encoding probes for a specific RNA species contain a unique combination of four of the N readout sequences, which correspond to the four hybridization rounds where this RNA should read '1'. N subsequent rounds of hybridization with the fluorescent readout probes were used to probe the readout sequences (hyb 1, hyb 2, ..., hyb N). The bound probes were inactivated by photobleaching between

successive rounds of hybridization. For clarity, only one possible pairing of the readout sequences is depicted here for the encoding probes; however, all possible pairs of the four readout sequences are used at the same frequency and distributed randomly along each cellular RNA in the actual experiments.

Fig. 11 shows a schematic description of a combinatorial labeling approach based on a simple binary code. In a conceptually simple labeling approach, 2^N-1 different RNA species can be uniquely encoded with all N-bit binary words (excluding the word with all '0's). In each hybridization round, FISH probes that are targeted to all RNA species that have a '1' in the corresponding bit are included. To increase the ability to discriminate RNA spots from background, each RNA is addressed with multiple FISH probes per hybridization round. Signal from the bound probes is extinguished before the next round of hybridization. This process continues for all N hybridization rounds (hyb 1, hyb 2, ...), and all 2^N-1 RNA species can be identified by the unique on-off pattern of fluorescence signals in each hybridization round.

Fig. 12 shows schematic descriptions of Hamming distance and its use in the identification and correction of errors. Fig. 12A is a schematic representation of a Hamming distance of 4. Figs. 12B and 12C are schematics showing the ability of an encoding scheme with Hamming distance 4 to correct single-bit errors (Fig. 12B) or detect but not correct double-bit errors (Fig. 12C). Arrows highlight bits at which the indicated words differ. Two code words are separated by a Hamming distance of 4 if one of the words has to flip four bits from '1' to '0' or '0' to '1' to convert into the other word. Single-bit error correction is possible because if a measured word differs from a legitimate code word by only one bit, it is most likely an error that arises from misreading this code word, since the code words of all the other RNA species will differ from the measured word by at least three bits. In this case, the measured word can be corrected to a code word that differs by only one bit. If a measured word differs from a legitimate code word by two bits, this measured word can still be identified as an error, but correction is no longer possible since more than one legitimate code word differs from this measured word by two bits.

Fig. 13 shows production of the library of encoding probes. An array-synthesized complex oligopool, containing ~100k sequences, is used as a template for the enzymatic amplification of the encoding probes for different experiments. Each template sequence in the oligopool contains a central target region that can bind to a cellular RNA, two flanking readout sequences, and two flanking index primers. In the first step, the required template molecules for a specific experiment are selected and amplified with an indexed PCR reaction.

To allow amplification via *in vitro* transcription, a T7 promotor is added to the PCR products during this step. In the second step, RNA is amplified from these template molecules via *in vitro* transcription. In the third step, this RNA is reverse transcribed back into DNA. In the final step, the template RNA is removed via alkaline hydrolysis, leaving only the desired ssDNA probes. This protocol produces ~2 nmol of complex pools of encoding probes containing ~20,000 different sequences for the 140-gene experiments or ~100,000 different sequences for the 1001-gene experiments.

EXAMPLE 5

This example illustrates the measurement of 140 genes with MERFISH using a 16-bit MHD4 Code. To test the feasibility of this error-robust, multiplexed imaging approach, this example uses a 140-gene measurement on human fibroblast cells (IMR90) using a 16-bit MHD4 code to encode 130 RNA species while leaving 10 code words as misidentification controls (Figs. 20A to 20H). After each round of hybridization with the fluorescent readout probes, cells were imaged by conventional wide-field imaging with an oblique-incidence illumination geometry. Fluorescent spots corresponding to individual RNAs were clearly detected and were then efficiently extinguished via a brief photobleaching step (Fig. 6A). The sample was stable throughout the 16 rounds of iterative labeling and imaging. The change in the number of fluorescent spots from round to round matched the predicted change based on the relative abundances of RNA species targeted in each round derived from bulk sequencing, and a systematic decreasing trend with increasing number of hybridization rounds was not observed (Fig. 14A). The average brightness of the spots varied from round to round with a standard deviation of 40%, likely due to different binding efficiencies of the readout probes to the different readout sequences on the encoding probes (Fig. 14B). Only a small, systematic decreasing trend in the spot brightness with increasing hybridization rounds was observed, which was on average 4% per round (Fig. 14B).

Next, binary words were constructed from the observed fluorescent spots based on their on-off patterns across the 16 hybridization rounds (Fig. 6B-6D). If the word exactly matched one of the 140 MHD4 code words (exact matches) or differed by only one bit (error-correctable matches), it was assigned to the corresponding RNA species (Fig. 6D). Within the single cell depicted in Figs. 6A and 6B, more than 1500 RNA molecules corresponding to 87% of the 130 encoded RNA species were detected after error correction (Fig. 6E). Similar observations were made in ~400 cells from 7 independent experiments. On average, ~4 times as many RNA molecules and ~2 times as many RNA species were detected per cell after error correction as compared with the values obtained before error correction (Fig. 15).

Two types of errors can occur in the copy number measurement of each RNA species: 1) Some molecules of this RNA species are not detected, leading to a drop in calling rate, and 2) some molecules from other RNA species are misidentified as this RNA species. To assess the extent of misidentification, the 10 misidentification control words were utilized, i.e., code words that were not associated with any cellular RNA. Although matches to these control words were observed, they occurred far less frequently than the real RNA-encoding words: 95% of the 130 RNA-encoding words were counted more frequently than the median count for these control words. Moreover, it was typically found that the ratio of the number of exact matches to the number of matches with one-bit errors for a real RNA-encoding word was substantially higher than the same ratios observed for the misidentification controls, as expected (Figs. 16A and 16B). Using this ratio as a measure of the confidence in RNA identification, it was found that 91% of the 130 RNA species had a confidence ratio greater than the maximum confidence ratio observed for the misidentification controls (Fig. 6F), demonstrating a high accuracy of RNA identification. Subsequent analyses were conducted only on these 91% of genes.

To estimate the calling rate, the error-correction ability of the MHD4 code was utilized to determine the 1-->0 error rates (10% on average) and 0-->1 error rates (4% on average) for each hybridization round (Figs. 16C and 16D). Using these error rates, an ~80% calling rate for individual RNA species after error correction was estimated, i.e. ~80% of the fluorescent spots corresponding to a RNA species were decoded correctly (Fig. 16E). It is noted that although the remaining 20% of spots contributed to a loss in detection efficiency, most of them did not cause species misidentification because they were decoded as double-bit error words and discarded.

To test for potential technical bias in these measurements, the same 130 RNAs species were probed with a different MHD4 codebook by shuffling the code words among different RNA species (Figs. 20A to 20H) and changing the encoding probe sequences. Measurements with this alternative code gave similar misidentification and calling rates (Fig. 17). The copy numbers of individual RNA species per cell measured with these two codebooks showed excellent agreement with a Pearson correlation coefficient of 0.94 (Fig. 6G), indicating that the choice of encoding scheme did not bias the measured counts.

In order to validate the copy numbers derived from the MERFISH experiments, conventional smFISH measurements were performed on 15 of the 130 genes, selected from the full measured abundance range of three orders of magnitude. For each of these genes, both the average copy number and the copy number distribution across many cells agreed

quantitatively between the MERFISH and conventional smFISH measurements (Fig. 18A and 18B). The ratio of the copy numbers determined by these two approaches was 0.82 +/- 0.06 (mean +/- SEM across the 15 measured RNA species, Fig. 18B), which agreed with the estimated 80% calling rate for the multiplexed imaging approach. The quantitative match between this ratio and the estimated calling rate over the full measured abundance range additionally supports the assessment that the misidentification error was low. Given that the agreement between the MERFISH and conventional smFISH results extended to the genes at the lowest measured abundance (<1 copy per cell, Fig. 18B), it was estimated that the measurement sensitivity was at least 1 copy per cell.

As a final validation, the abundance of each RNA species averaged over hundreds of cells was compared to those obtained from a bulk RNA sequencing measurement that were performed on the same cell line. The imaging results correlated remarkably well with bulk sequencing results with a Pearson correlation coefficient of 0.89 (Fig. 6H).

Fig. 6 shows simultaneous measurement of 140 RNA species in single cells using MERFISH with a 16-bit MHD4 code. Fig. 6A shows images of RNA molecules in an IMR90 cell after each hybridization round (hyb 1 – hyb 16). The image after photobleaching (bleach 1) demonstrated efficient removal of fluorescent signals between hybridizations. Fig. 6B shows the localizations of all detected single molecules in this cell colored based on their measured binary words. Inset: the composite fluorescent image of the 16 hybridization rounds for the boxed sub-region with numbered circles indicating potential RNA molecules. A circle indicates an unidentifiable molecule, the binary word of which does not match any of the 16-bit MHD4 code words even after error correction. Fig. 6C shows fluorescent images from each round of hybridization for the boxed sub-region in Fig. 6B with circles indicating potential RNA molecules. Fig. 6D shows corresponding words for the spots identified in Fig. 6C. Crosses represent the corrected bits. Fig. 6E shows the RNA copy number for each gene observed without (lower) or with (higher) error correction in this cell. Fig. 6F shows the confidence ratio measured for the 130 RNA species (left) and the 10 misidentification control words (right) normalized to the maximum value observed from the misidentification controls (dashed line). Fig. 6G is a scatter plot of the average copy number of each RNA species per cell measured with two shuffled codebooks of the MHD4 code. The Pearson correlation coefficient is 0.94 with a p-value of 1×10^{-53}. The dashed line corresponds to the $y = x$ line. Fig. 6H is a scatter plot of the average copy number of each RNA species per cell versus the abundance determined by bulk sequencing in fragments per kilobase per million reads

(FPKM). The Pearson correlation coefficient between the logarithmic abundances of the two measurements was 0.89 with a p-value of $3x10^{-39}$.

Fig. 14 shows the number and average brightness of the fluorescent spots detected in the 16 rounds of hybridization before and after photobleaching. Fig. 14A shows the number of fluorescent spots observed per cell before (higher) and after (lower) photobleaching as a function of hybridization round averaged across all measurements with the first 16-bit MHD4 code. Photobleaching reduces the number of fluorescent spots by two or more orders of magnitude. Hybridization rounds without lower bars represent rounds in which no molecules were observed after bleaching. Also depicted is the expected change in the number of fluorescent spots from round to round (circles) predicted based on the relative abundances of the RNA species targeted in each hybridization round derived from bulk RNA sequencing. The average discrepancy between the observed and predicted number of spots for each hybridization is only 15% of the mean number of spots. This discrepancy does not systematically increase with the number of hybridization rounds. Fig. 14B shows the average brightness of the identified fluorescent spots in each hybridization round averaged across all measurements with the first 16-bit MHD4 code both before (upper) and after (lower) photobleaching. Brightness varies by 40% (standard deviation) across different hybridization rounds. The variation pattern is reproducible between experiments with the same code, likely due to differences in the binding efficiency of the readout probes to the different readout sequences. There is a small systematic trend of decrease in the brightness with increasing hybridization rounds, which is on average 4% per round. Photobleaching extinguishes fluorescence to a level similar to that of the autofluorescence of the cell.

Fig. 15 shows error correction substantially increases the numbers of RNA molecules and RNA species detected in individual cells. Fig. 15A shows a histogram of the ratio of the total number of molecules detected per cell with error correction to the number measured without error correction. Fig. 15B is a histogram of the total number of RNA species detected in each cell with error correction to that without error correction. Both ratios are determined for ~200 cells and the histograms are constructed from these ratios.

Fig. 16 shows characterization of the misidentification and calling rates of RNA species for the 140-gene experiments using a specific 16-bit MHD4 code. Fig. 16A shows the number of measured words exactly matching the code word corresponding to FLNA, represented by the bar in the center of the circle, and the number of measured words with one-bit error compared to the code word of FLNA, represented by the 16 bars on the circle. Fig. 16B is the same as Fig. 16A, but for a code word that was not assigned to any RNA, i.e.,

a misidentification control word. The solid lines connect the exact match to one-bit error words that are generated by 1-->0 errors. Based on the observation that the ratio of the number of exact matches to the number of error-correctable matches for a real RNA-encoding word was typically substantially higher than the same ratios observed for the misidentification controls, this ratio was defined as a confidence ratio for RNA identification. The confidence ratio measured for all 130 RNA species (center bar) and 10 misidentification control words not assigned to any RNA (outer bars) using this 16-bit MHD4 code is show in Fig. 6F. Fig. 16C and 16D show the average error rates for the 1-->0 error (Fig. 16C) and 0-->1 error (Fig. 16D) for each hybridization round. Fig. 16E shows the calling rate for each RNA species estimated from the 1-->0 and 0-->1 error rates. Genes are sorted from left to right based on the measured abundance, which spans three orders of magnitude. The calling rates are largely independent of the abundance of the gene.

Fig. 17 shows characterization of the misidentification and calling rates for a second 16-bit MHD4 code. In this second encoding scheme, the 140 code words were shuffled among different RNA species and changed the encoding probe sequences. Fig. 17A shows the normalized confidence ratio measured for the 130 RNA species (left) and the 10 misidentification control words not assigned to any RNA (right). The normalized confidence ratio is determined the same way as in Fig. 6F. Figs. 17B and 17C show the average error rates determined for the 1-->0 error (Fig. 17B) and 0-->1 error (Fig. 17C) for each hybridization round. Fig. 17D shows the calling rate determined for each RNA species estimated from the 1-->0 and 0-->1 error rates. Genes are sorted from left to right based on the measured abundance.

Fig. 18 shows a comparison of the MERFISH measurements with conventional smFISH results for a subset of genes. Fig. 18A shows the distributions of RNA copy numbers in single cells for three example genes KIAA1199, DYNC1H1, and LMTK2 in the high, medium, and low abundance ranges, respectively. Lighter bars: distributions constructed from ~400 cells in the 140-gene measurements using the MHD4 codes. Darker bars: distributions constructed from ~100 cells in the conventional smFISH measurements. Fig. 18B shows a comparison of the average RNA copy numbers per cell measured in the 140-gene experiments using the MHD4 codes to those determined by conventional smFISH for 15 genes. The average ratio of the copy number measured using the MHD4 measurements to that measured using conventional smFISH was 0.82 +/- 0.06 (mean +/- SEM across 15 genes). The dashed line corresponds to the $y = x$ line.

Figs. 20A to 20H show two different codebooks for the 140-gene experiments. The specific code words of the 16-bit MHD4 code assigned to each RNA species studied in the two shuffles of the 140-gene experiment. The "Genes" columns contain the name of the gene. The "Codewords" columns contain the specific binary word assigned to each gene.

EXAMPLE 6

This example is generally directed to high-throughput analysis of cell-to-cell variation in gene expression. The MERFISH approach allows parallelization of measurements of many individual RNA species and co-variation analysis between different RNA species. In this example, the parallelization aspect was first illustrated by examining the cell-to-cell variation in the expression level of each of the measured genes (Fig. 7A). To quantify the measured variation, the Fano factors, defined as the ratio of the variance to the mean RNA copy number, were computed for all measured RNA species. The Fano factors substantially deviated from 1, the value expected for a simple Poisson process, for many genes and exhibited an increasing trend with the mean RNA abundance (Fig. 7B). This trend of increasing Fano factors with mean RNA abundance can be explained by changes in the transcription rate and/or promoter off-switching rates but not by changes in the promoter on-switching rate.

Moreover, several RNA species were identified with substantially larger Fano factors than this average trend. For example, it was found that SLC5A3, CENPF, MKI67, TNC and KIAA1199 displayed Fano factor values substantially higher than those of the other genes expressed at similar abundance levels. The high variability of some of these genes can be explained by their association with the cell cycle. For example, two of these particularly 'noisy' genes MKI67 and CENPF were both annotated as cell-cycle related genes, and based on their bimodal expression (Fig. 7C), it is proposed that their transcription is strongly regulated by the cell cycle. Other high-variability genes did not show the same bimodal expression patterns and are not known to be associated with the cell cycle.

Analysis of co-variations in the expression levels of different genes can reveal which genes are co-regulated and elucidate gene regulatory pathways. At the population level, such analysis often requires the application of external stimuli to drive gene expression variation; hence, correlated expression changes can be observed among genes that share common regulatory elements influenced by the stimuli. At the single-cell level, one can take advantage of the natural stochastic fluctuations in gene expression for such analysis and can thus study multiple regulatory networks without having to stimulate each of them

individually. Such co-variation analysis can constrain regulatory networks, suggest new regulatory pathways, and predict function for unannotated genes based on associations with co-varying genes.

This approach was applied to the 140-gene measurements and the ~10,000 pairwise correlation coefficients describing how the expression levels of each pair of genes co-varied from cell to cell were examined. Many of the highly variable genes showed tightly correlated or anti-correlated variations (Fig. 7C). To better understand the correlations for all gene pairs, a hierarchical clustering approach was adopted to organize these genes based on their correlation coefficients (Fig. 7D). From the cluster tree structure, seven groups of genes with substantially correlated expression patterns were identified (Fig. 7D). Within each of the seven groups, every gene showed significantly stronger average correlation with other members of the group than with genes outside the group. To further validate and understand these groups, gene ontology (GO) terms enriched in each of these seven groups were identified. Notably, the enriched GO terms within each group shared similar functions and were largely unique to each group (Fig. 7E), validating the notion that the observed co-variation in expression reflects some commonalities in the regulation of these genes.

This example describes two of these groups as illustrative examples. The predominant GO terms associated with Group 1 were terms associated with the extracellular matrix (ECM) (Figs. 7D and 7E). Notable members of this group included ECM components, such as FBN1, FBN2, COL5A, COL7A and TNC, and glycoproteins linking the ECM and cell membranes, such as VCAN and THBS1. The group also included an unannotated gene, KIAA1199, which may predict to play a role in ECM metabolism based on its association with this cluster. Indeed, this gene has recently been identified as an enzyme involved in the regulation of hyaluronan, a major sugar component of the ECM.

Group 6 contained many genes that encode vesicle transport proteins and proteins associated with cell motility (Figs. 7D and 7E). The vesicle transport genes included microtubule motors and related genes DYNC1H, CKAP1, and factors associated with vesicle formation and trafficking, like DNAJC13 and RAB3B. Again, an unannotated gene, KIAA1462, was found within this cluster. Based on its strong correlation with DYNC1H1 and DNAJC13, it is predicted that this gene may be involved in vesicle transport. The cell motility genes in this group included actin-binding proteins like AFAP1, SPTAN1, SPTBN1, and MYH10, and genes involved in the formation of adhesion complexes, like FLNA and FLNC. Several GTPase-associated factors involved in the regulation of cell motility, attachment and contraction also fell into this group, including DOCK7, ROCK2, IQGAP1,

PRKCA, and AMOTL1. The observation that some cell motility genes correlated with vesicle transport genes is consistent with the role of vesicle transport in cell migration. An additional interesting feature of group 6 is that a subset of these genes, in particular those related to cell motility, were anti-correlated with members of the ECM group discussed above (Fig. 7D). This anti-correlation may reflect regulatory interactions that mediate switching of cells between adherent and migratory states.

Fig. 7 shows cell-to-cell variations and pairwise correlations for the RNA species determined from the 140 gene measurements. Fig. 7A shows a comparison of gene expression levels in two individual cells. Fig. 7B shows Fano factors for individual genes. Error bars represent standard error of the mean determined from 7 independent data sets. Fig. 7C shows Z-scores of the expression variations of four example pairs of genes showing correlated (top two) or anti-correlated (bottom two) variation for 100 randomly selected cells. The Z-score is defined as the difference from the mean normalized by the standard deviation. Fig. 7D is a matrix of the pairwise correlation coefficients of the cell-to-cell variation in expression for the measured genes, shown together with the hierarchical clustering tree. The seven groups identified by a specific threshold on the cluster tree (dashed line) are indicated by the black boxes in the matrix and lines on the tree, with grey lines on the tree indicating ungrouped genes. Different threshold choices on the cluster tree could be made to select either smaller subgroups with tighter correlations or larger super-groups containing more weakly coupled subgroups. Two of the seven groups are enlarged on the right. Fig. 7E shows enrichment of 30 selected, statistically significantly enriched GO terms in the seven groups. Enrichment refers to the ratio of the fraction of genes within a group that have the specific GO term to the fraction of all measured genes having that term. Not all of the GO terms presented here are in the top 10 list.

EXAMPLE 7

This example illustrates mapping spatial distributions of RNAs. As an imaging based approach, MERFISH also allowed the investigation of the spatial distributions of many RNA species simultaneously. Several patterns emerged from the visual inspection of individual genes, with some RNA transcripts enriched in the perinuclear region, some enriched in the cell periphery, and some scattered throughout the cell (Fig. 8A). To identify genes with similar spatial distributions, the correlation coefficients for the spatial density profiles of all pairs of RNA species were determined, and these RNAs were organized based on the pairwise correlations again using the hierarchical clustering approach. The correlation coefficient matrix showed groups of genes with correlated spatial organizations, and the two

most notable groups with the strongest correlations are indicated in Fig. 8B. Group I RNAs appeared enriched in the perinuclear region whereas group II RNAs appeared enriched near the cell peripheral region (Fig. 8C). Quantitative analysis of the distances between each RNA molecule and the cell nucleus or the cell periphery indeed confirmed this visual impression (Fig. 8D).

Group I contained genes encoding extracellular proteins such as FBN1, FBN2 and THSB1, secreted proteins such as PAPPA, and integral membrane proteins such as LRP1 and GPR107. These proteins have no obvious commonalities in function. Rather a GO analysis showed significant enrichment for location terms, such as extracellular region, basement membrane, or perivitelline space (Fig. 8E). To reach these locations, proteins must pass through the secretion pathway, which often requires translation of mRNA at the endoplasmic reticulum (ER). Thus, it is proposed that the spatial pattern that were observed for these mRNAs reflects their co-translational enrichment at the ER. The enrichment of these mRNAs in the perinuclear region (Figs. 8C and 8D, lighter shading), where the rough ER resides, supports this conclusion.

Group II contained genes encoding the actin-binding proteins, including filamins FLNA and FLNC, talin TLN1, and spectrins SPTAN1 and SPTBN1; the microtubule-binding protein CKAP5; and the motor proteins MYH10 and DYNC1H1. This group was enriched with GO terms such as cortical actin cytoskeleton, actin filament binding, and cell-cell adherens junction (Fig. 8E). Beta-actin mRNA may be enriched near the cell periphery in fibroblasts as are mRNAs that encode members of the actin-binding Arp2/3 complex. The enrichment of group II mRNAs in the peripheral region of the cells (Figs. 8C and 8D) suggests that the spatial distribution of the Group II genes might be related to the distribution of actin cytoskeleton mRNAs.

Fig. 8 shows distinct spatial distributions of RNAs observed in the 140-gene measurements. Fig. 8A shows examples of the spatial distributions observed for four different RNA species in a cell. Fig. 8B is a matrix of the pairwise correlation coefficients describing the degree with which the spatial distributions of each gene pair is correlated, shown together with the hierarchical clustering tree. Two strongly correlating groups are indicated by the black boxes on the matrix and shading on the tree. Fig. 8C shows the spatial distributions of all RNAs in the two groups in two example cells. Lighter symbols: group I genes; darker symbols: group II genes. Fig. 8D shows average distances for genes in group I and genes in group II to the cell edge or the nucleus normalized to the average distances for

all genes. Error bars represent SEM across 7 data sets. Fig. 8E shows enrichment of GO terms in each of the two groups.

EXAMPLE 8

This example illustrates measuring 1001 genes with a 14-bit MHD2 code. This example further increases the throughput of MERFISH measurements by simultaneously imaging ~1000 RNA species. This increase could be achieved with the MHD4 code by increasing the number of bits per code word to 32 while maintaining the number of '1' bits per word at four (Fig. 5B). While the stability of the samples across many hybridization rounds (Fig. 14) suggests that such an extension is potentially feasible, an alternative approach is shown here that did not require an increase in the number of hybridizations by relaxing the error correction requirement but keeping the error detection capability. For example, by reducing the Hamming distance from 4 to 2, all 14-bit words could be used that contain four '1' bits to encode 1001 genes and these RNAs were probed with only 14 rounds of hybridization. However, because a single error can produce a word equally close to two different code words, error correction is no longer possible for this modified Hamming-distance-2 (MHD2) code. Hence, it was expected that the calling rate would be lower and the misidentification rate to be higher with this encoding scheme.

To evaluate the performance of this 14-bit MHD2 code, 16 of the 1001 possible code words were set aside as misidentification controls and used the remaining 985 words to encode cellular RNAs. Among these 985 RNAs included 107 RNA species probed in the 140-gene experiments as an additional control. The 1001-gene experiments were performed in IMR90 cells using a similar procedure as described above. To allow all encoding probes to be synthesized from a single 100,000-member oligopool, the number of encoding probes per RNA species was reduced to ~94. Fluorescent spots corresponding to individual RNA molecules were again detected in each round of hybridization with the readout probes and, based on their on-off patterns, these spots were decoded into RNA (Figs. 9A, 19A and 19B). 430 RNA species were detected in the cell shown in Fig. 9A, and similar results were obtained in ~200 imaged cells in 3 independent experiments.

As expected, the misidentification rate of this scheme was higher than that of the MHD4 code. 77% of all real RNA words were detected more frequently than the median count for the misidentification controls instead of the 95% value observed in the MHD4 measurements. Using the same confidence ratio analysis as described above, it was found that 73% (instead of 91% for the MHD4 measurements) of the 985 RNA species were measured with a confidence ratio larger than the maximum value observed for the

misidentification controls (Fig. 19C). RNA copy numbers measured from these 73% RNA species showed excellent correlation with the bulk RNA sequencing results (Pearson correlation coefficient r = 0.76; Fig. 9B, black). It is worth noting that the remaining 27% of the genes still exhibit good, albeit lower, correlation with the bulk RNA sequencing data (r = 0.65; Fig. 9B, red), but the conservative measure of excluding them from further analysis was taken.

The lack of an error correction capability also decreased the calling rate of each RNA species: When comparing the 107 RNA species common in both the 1001-gene and 140-gene measurements, it was found that the copy numbers per cell of these RNA species were lower in the 1001-gene measurements (Figs. 9C and 19D). The total count of these RNAs per cell was ~1/3 of that observed in the 140-gene measurements. Thus the lack of error correction in the MHD2 code produced a ~3-fold decrease in the calling rate, which is consistent with the ~4-fold decrease in calling rate observed for the MHD4 code when error correction was not applied. As expected from the quantitative agreement between 140-gene measurements and conventional smFISH results, comparison of the 1001-gene measurements with conventional smFISH results for 10 RNA species also indicated a ~3-fold drop in calling rate (Fig. 18C). Despite the expected reduction in calling rate, the good correlations found between the copy numbers observed in the 1001-gene measurements and those observed in the 140-gene measurements, as well as in conventional smFISH and bulk RNA sequencing measurements, indicates that the relative abundance of these RNAs can be quantified with the MHD2 encoding scheme.

Simultaneously imaging ~1000 genes in individual cells substantially expanded the ability to detect co-regulated genes. Fig. 10A shows the matrix of pairwise correlation coefficients determined from the cell-to-cell variations in the expression levels of these genes. Using the same hierarchical clustering analysis as described above, ~100 groups of genes with correlated expression were identified. Remarkably, nearly all of these ~100 groups showed statistically significant enrichment of functionally related GO terms (Fig. 10B). These included some of the groups identified in the 140-gene measurements, such as the group associated with cell replication genes and the group associated with cell motility genes (Figs. 10A and 10B, groups 7 and 102), as well as many new groups. The groups identified here included 46 RNA species lacking any previous GO annotations, for which function based on their group association may be hypothesized. For example, KIAA1462 is part of the cell motility group, as also shown in the 140-gene experiments, suggesting a potential role of this gene in cell motility (Fig. 10A, group 102). Likewise, KIAA0355 is part

of a new group enriched in genes associated with heart development (Fig. 10A, group 79), and C17orf70 is part of a group associated with ribosomal RNA processing (Fig. 10A, group 22). Using these groupings, cellular functions for 61 transcription factors and other partially annotated proteins of unknown functions may be hypothesized. For example, the transcription factors Z3CH13 and CHD8 are both members of the cell motility group, suggesting their potential role in the transcriptional regulation of cell motility genes.

Fig. 9 shows simultaneous measurements of 1001 genes in single cells using MERFISH with a 14-bit MHD2 code. Fig. 9A shows the localizations of all detected single molecules in a cell colored based on their measured binary words. Inset: the composite, false-colored fluorescent image of the 14 hybridization rounds for the boxed sub-region with numbered circles indicating potential RNA molecules. Circles indicate unidentifiable molecules, the binary words of which do not match any of the 14-bit MHD2 code words. Images of individual hybridization round are shown in Fig. 19A. Fig. 9B is a scatter plot of the average copy number per cell measured in the 1001-gene experiments versus the abundance measured via bulk sequencing. The upper symbols are for the 73% of genes detected with confidence ratios higher than the maximum ratio observed for the misidentification controls. The Pearson correlation coefficient is 0.76 with a p-value of 3×10^{-133}. The lower symbols are for the remaining 27% of genes. The Pearson correlation coefficient is 0.65 with a p-value of 3×10^{-33}. Fig. 9C is a scatter plot of the average copy number for the 107 genes shared in both the 1001-gene measurement with the MHD2 code and the 140-gene measurement with the MHD4 code. The Pearson correlation coefficient is 0.89 with a p-value of 9×10^{-30}. The dashed line corresponds to the $y = x$ line.

Fig. 10 shows co-variation analysis of the RNA species measured in the 1001-gene measurements. Fig. 10A is a matrix of all pairwise correlation coefficients of the cell-to-cell variation in expression for the measured genes shown with the hierarchical clustering tree. The ~100 identified groups of correlated genes are indicated by shading on the tree. Zoom in of four of the groups described in the text are shown on the right. Fig. 10B is an enrichment of 20 selected, statistically significantly enriched GO terms in the four groups.

Fig. 19 shows decoding and error assessment of the 1001-gene experiments. Fig. 19A shows images of the boxed sub-region of the cell in Fig. 9A for each of 14 hybridization rounds. The final panel is a composite image of these 14 rounds. Circles indicate fluorescent spots that have been identified as potential RNA molecules. Some circles in the composite image indicate unidentifiable molecules, the binary words of which do not match any of the 14-bit MHD2 code words. Fig. 9B shows the corresponding binary word for each of the

spots identified in Fig. 9A with the RNA species to which it is decoded. 'unidentified' implies that the measured binary word does not match any of the 1001 code words. Fig. 19C shows the normalized confidence ratios measured for the 985 RNA species (left) and the 16 misidentification control words not targeted to any RNA (right). The normalized confidence ratio is defined as in Fig. 6F. Fig. 19D shows a histogram of the reduction in detected abundance of 107 genes present in both the 1001-gene experiments and the 140-gene experiments. "Fold decrease in copy number" is defined as the average number of RNA molecules per cell for each species measured in the 140-gene experiments divided by the corresponding average number measured in the 1001-gene experiments.

Fig. 18C is a comparison of the average RNA copy numbers per cell measured in the 1001-gene experiments using the MHD2 code to those determined by conventional smFISH for 10 genes. The average ratio of the copy number measured using the MHD2 measurements to that measured using conventional smFISH was 0.30 +/- 0.05 (mean +/- SEM across 10 genes). The dashed line corresponds to the $y = x$ line and the dotted line corresponds to the $y = 0.30x$ line.

EXAMPLE 9

The above examples illustrate a highly multiplexed detection scheme for systems-level RNA imaging in single cells. Using combinatorial labeling, sequential hybridization and imaging, and two different error-robust encoding schemes, either 140 or 1001 genes in hundreds of individual human fibroblast cells were simultaneously imaged. Of the two encoding schemes presented here, the MHD4 code is capable of both error detection and error correction, and hence can provide a higher calling rate and a lower misidentification rate than the MHD2 code, which instead can only detect but cannot correct errors. MHD2, on the other hand, provides a faster scaling of the degree of multiplexing with the number of bits than MHD4. Other error-robust encoding schemes can also be used for such multiplexed imaging, and experimenters can set the balance between detection accuracy and ease of multiplexing based on the specific requirements of the experiments.

By increasing the number of bits in the code words, it should be possible to further increase the number of detectable RNA species using MERFISH with, for example, a MHD4 or MHD2 code. For example, using the MHD4 code with 32 total bits and four or six '1' bits would increase the number of addressable RNA species to 1,240 or 27,776, respectively. The latter is the approximate scale of the human transcriptome. The predicted misidentification and calling rates are still reasonable for the 32-bit MHD4 code (shown in Figs. 5C and 5D for the MHD4 code with four '1' bits and similar rates were calculated for the MHD4 code with

six '1' bits). If more accurate measurements are desired, an additional increase in the number of bits would allow the use of encoding schemes with a Hamming distance greater than 4, further enhancing the error detection and correction capability. While an increase in the number of bits by adding more hybridization rounds would increase the data collection time and potentially lead to sample degradation, these problems could be mitigated by utilizing multiple colors to readout multiple bits in each round of hybridization.

As the degree of multiplexing is increased, it is important to consider the potential increase in the density of RNAs that need to be resolved in each round of imaging. Based on the imaging and sequencing results, it can be estimated that including the whole transcriptome of the IMR90 cells would lead to a total RNA density of ~200 molecules/micrometer^3. Using the current imaging and analysis methods, 2-3 molecules/micrometer^3 per hybridization round could be resolved, which would reach a total RNA density of ~20 molecules/micrometer^3 after 32 rounds of hybridization. This density should allow all but the top 10% most expressed genes to be imaged simultaneously or a subset of genes with even higher expression levels to be included. By utilizing more advanced image analysis algorithms to better resolve overlapping images of individual molecules, such as compressed sensing, it is possible to extend the resolvable density by ~4-fold and thus allow all but the top 2% most expressed genes to be imaged all together.

These examples have illustrated the utility of the data derived from highly multiplexed RNA imaging by using co-variation and correlation analysis to reveal distinct sub-cellular distribution patterns of RNAs, to constrain gene regulatory networks, and to predict functions for many previously unannotated or partially annotated genes with unknown functions. Given its ability to quantify RNAs across a wide range of abundances without amplification bias while preserving native context, systems and methods such as MERFISH will allow many applications of *in situ* transcriptomic analyses of individual cells in culture or complex tissues.

EXAMPLE 10

Following are various materials and methods used in the above examples.

Probe design. Each RNA species in the target set was randomly assigned a binary code word either from all 140 possible code words of the 16-bit MHD4 code or from all 1001 possible code words of the 14-bit MHD2 code.

Array-synthesized oligopools were used as templates to make the encoding probes. The template molecule for each encoding probe contained three components: i) a central targeting sequence for *in situ* hybridization to the target RNA, ii) two flanking readout

sequences designed to hybridize each of two distinct readout probes, and iii) two flanking primer sequences to allow enzymatic amplification of the probes (Fig. 13). The readout sequences were taken from the 16 possible readout sequences, each corresponding to one hybridization round. The readout sequences were assigned to the encoding probes such that for any RNA species each of the 4 readout sequences were distributed uniformly along the length of the target RNA and appeared at the same frequency. Template molecules for the 140-gene library also included a common 20-nucleotide (nt) priming region between the first PCR primer and the first readout sequence. This priming sequence was used for the reverse transcription step described below.

Multiple experiments were embedded in a single array-synthesized oligopool, and PCR was used to selectively amplify only the oligos required for a specific experiment. Primer sequences for this indexed PCR reaction were generated from a set of orthogonal 25-nt sequences. These sequences were trimmed to 20 nt and selected for i) a narrow melting temperature range (70 °C to 80 °C), ii) the absence of consecutive repeats of 3 or more identical nucleotides, and iii) the presence of a GC clamp, i.e. one of the two 3' terminal bases must be G or C. To further improve specificity, these sequences were then screened against the human transcriptome using BLAST+, and primers with 14 or more contiguous bases of homology were eliminated. Finally, BLAST+ was again used to identify and exclude primers that had an 11-nt homology region at the 3' end of any other primer or a 5-nt homology region at the 3' end of the T7 promoter. The forward primer sequences (Primer 1) were determined as described above, whereas the reverse primers each contain a 20-nt sequence as described above plus a 20-nt T7 promoter sequence to facilitate amplification via *in vitro* transcription (Primer 2). The primer sequences used in the 140-gene and 1001-gene experiments are listed below.

Table 2

Experiment Name	Primer 1 Sequence (Index Primer 1)	Primer 2 Sequence (T7 promoter plus the reverse complement of Index Primer 2)
140-gene Codebook 1	GTTGGTCGGCACTTGGGTGC (SEQ ID NO: 18)	TAATACGACTCACTATAGGGAAAGCCGGTTCATCCGGTGG (SEQ ID NO: 21)
140-gene Codebook 2	CGATGCGCCAATTCCGGTTC (SEQ ID NO: 19)	TAATACGACTCACTATAGGGTGATCATCGCTCGCGGGTTG (SEQ ID NO: 22)
1001-gene	CGCGGGCTATATGCGAACC	TAATACGACTCACTATAGGGCGTGGAGG

G (SEQ ID NO: 20) GCATACAACGC (SEQ ID NO: 23)

30-nt-long readout sequences were created by concatenating fragments of the same orthogonal primer set generated above by combining one 20-nt primer with a 10-nt fragment of another. These readout sequences were then screened, using BLAST+, for orthogonality with the index primer sequences and other readout sequences (no more than 11 nt of homology) and for potential off-target binding sites in the human genome (no more than 14 nt of homology). Fluorescently labeled readout probes with sequences complementary to the readout sequences were used to probe these readout sequences, one in each hybridization round. All used readout probes sequences are listed below.

Table 3

Bit	Readout probes	
1	CGCAACGCTTGGGACGGTTCCAATCGGATC/3Cy5Sp/	SEQ ID NO: 24
2	CGAATGCTCTGGCCTCGAACGAACGATAGC/3Cy5Sp/	SEQ ID NO: 25
3	ACAAATCCGACCAGATCGGACGATCATGGG/3Cy5Sp/	SEQ ID NO: 26
4	CAAGTATGCAGCGCGATTGACCGTCTCGTT/3Cy5Sp/	SEQ ID NO: 27
5	GCGGGAAGCACGTGGATTAGGGCATCGACC/3Cy5Sp/	SEQ ID NO: 28
6	AAGTCGTACGCCGATGCGCAGCAATTCACT/3Cy5Sp/	SEQ ID NO: 29
7	CGAAACATCGGCCACGGTCCCGTTGAACTT/3Cy5Sp/	SEQ ID NO: 30
8	ACGAATCCACCGTCCAGCGCGTCAAACAGA/3Cy5Sp/	SEQ ID NO: 31
9	CGCGAAATCCCCGTAACGAGCGTCCCTTGC/3Cy5Sp/	SEQ ID NO: 32
10	GCATGAGTTGCCTGGCGTTGCGACGACTAA/3Cy5Sp/	SEQ ID NO: 33
11	CCGTCGTCTCCGGTCCACCGTTGCGCTTAC/3Cy5Sp/	SEQ ID NO: 34
12	GGCCAATGGCCCAGGTCCGTCACGCAATTT/3Cy5Sp/	SEQ ID NO: 35
13	TTGATCGAATCGGAGCGTAGCGGAATCTGC/3Cy5Sp/	SEQ ID NO: 36
14	CGCGCGGATCCGCTTGTCGGGAACGGATAC/3Cy5Sp/	SEQ ID NO: 37
15	GCCTCGATTACGACGGATGTAATTCGGCCG/3Cy5Sp/	SEQ ID NO: 38
16	GCCCGTATTCCCGCTTGCGAGTAGGGCAAT/3Cy5Sp/	SEQ ID NO: 39

The readout probes used for the 140-gene libraries were probes 1 through 16. The readout probes used for the 1001-gene experiment were probes 1 through 14. /3Cy5Sp/ indicates a 3' Cy5 modification.

To design the central targeting sequences of the encoding probes, the abundance of different transcripts in IMR90 cells using Cufflinks v2.1, total RNA data from the ENCODE project, and human genome annotations from Gencode v18 were complied. Probes were designed from gene models corresponding to the most abundant isoform using OligoArray2.1 with the following constraints: the target sequence region is 30-nt long; the melting temperatures of the hybridized region of the probe and cellular RNA target is greater than 70 °C; there is no cross hybridization targets with melting temperatures greater than 72 °C; there

is no predicted internal secondary structures with melting temperatures greater than 76 °C; and there is no contiguous repeats of 6 or more identical nucleotides. Melting temperatures were adjusted to optimize the specificity of these probes and minimize secondary structure while still producing sufficient numbers of probes for the libraries. To decrease computational cost, isoforms were divided into 1-kb regions for probe design. Using BLAST+, all potential probes that mapped to more than one cellular RNA species were rejected. Probes with multiple targets on the same RNA were kept.

For each gene in the 140-gene experiments, 198 putative encoding probe sequences were generated by concatenating the appropriate index primers, readout sequences, and targeting regions as shown in Fig. 13. To address the possibility that concatenation of these sequences introduced new regions of homology to off-target RNAs, BLAST+ was used to screen these putative sequences against all human rRNA and tRNA sequences as well as highly expressed genes (genes with FPKM > 10,000). Probes with greater than 14 nt of homology to rRNAs or tRNAs or greater than 17 nt of homology to highly expressed genes were removed. After these cuts, there were ~192 (with a standard deviation of 2) probes per gene for both MHD4 codebooks used in the 140-gene experiments. The same protocol for the 1001-gene experiments was used, as follows: Starting with 96 putative targeting sequences per gene, ~94 (with a standard deviation of 6) encoding probes per gene were obtained after these additional homology cuts. The number of encoding probes per RNA was decreased for the 1001-gene experiments so that these probes could be synthesized from a single 100,000-member oligopool as opposed to two separate pools. Each encoding probe was designed to contain two of the four readout sequences associated with each code word, hence only half of the bound encoding probes can bind readout probe during any given hybridization round. ~192 or ~94 encoding probes per RNA were used to obtain high signal-to-background ratios for individual RNA molecules. The number of encoding probes per RNA could be substantially reduced but still allow single RNA molecules to be identified. In addition, increasing the number of readout sequences per encoding probe or using optical sectioning methods to reduce the fluorescence background may allow further reduction in the number of the encoding probes per RNA.

Two types of misidentification controls were designed. The first control (blank words) were not represented with encoding probes. The second type of control (no-target words) had encoding probes that were not targeting any cellular RNA. The targeting regions of these probes were composed of random nucleotide sequences subject to the same constraints used to design the RNA targeting sequences described above. Moreover, these

random sequences were screened against the human transcriptome to ensure that they contain no significant homology (>14-nt) to any human RNA. The 140-gene measurements contained 5 blank words and 5 no-target words. The 1001-gene measurements contained 11 blank words and 5 no-target words.

Probe synthesis. The encoding probes were synthesized using the following steps, and this synthesis protocol is illustrated in Fig. 13.

Step 1: The template oligopool (CustomArray) was amplified via limited-cycle PCR on a Bio-Rad CFX96 using primer sequences specific to the desired probe set. To facilitate subsequent amplification via *in vitro* transcription, the reverse primer contained the T7 promoter. All primers were synthesized by IDT. This reaction was column purified (Zymo DNA Clean and Concentrator; D4003).

Step 2: The purified PCR products were then further amplified ~200-fold and converted into RNA via a high yield *in vitro* transcription according to the manufacturer's instructions (New England Biolabs, E2040S). Each 20 microliter reaction contained ~1 microgram of template DNA from above, 10 mM of each NTP, 1x reaction buffer, 1x RNase inhibitor (Promega RNasin, N2611) and 2 microliters of the T7 polymerase. This reaction was incubated at 37 °C for 4 hours to maximize yield. This reaction was not purified before the following steps.

Step 3: The RNA products from the above *in vitro* transcription reaction were then converted back into DNA via a reverse transcription reaction. Each 50 microliter reaction contained the unpurified RNA produce from Step 2 supplemented with 1.6 mM of each dNTP, 2 nmol of a reverse transcription primer, 300 units of Maxima H- reverse transcriptase (Thermo Scientific, EP0751), 60 units of RNasin, and a final 1x concentration of the Maxima RT buffer. This reaction was incubated at 50 °C for 45 minutes, and the reverse transcriptase was inactivated at 85 °C for 5 minutes. The templates for the 140-gene libraries contain a common priming region for this reverse transcription step; thus, a single primer was used for this step when creating these probes. Its sequence was CGGGTTTAGCGCCGGAAATG (SEQ ID NO: 40). A common priming region was not included for the 1001-gene library; thus, the reverse transcription was conducted with the forward primer: CGCGGGCTATATGCGAACCG (SEQ ID NO: 20).

Step 4: To remove the template RNA, 20 microliters of 0.25 M EDTA and 0.5 N NaOH was added to the above reaction to selectively hydrolyze RNA, and the sample was incubated at 95 °C for 10 minutes. This reaction was then immediately purified by column purification using a 100-microgram-capacity column (Zymo Research, D4030) and the Zymo

Oligo Clean and Concentrator protocol. The final probes were eluted in 100 microliters of RNase-free deionized water, evaporated in a vacuum concentrator, and then resuspended in 10 μL of encoding hybridization buffer (see below). Probes were stored at -20 °C. Denaturing poly-acrylamide gel electrophoresis and absorption spectroscopy were used to confirm the quality of the probes and revealed that this probe synthesis protocol converts 90-100% of the reverse-transcription primer into full length probe and of the probe that is constructed, 70-80% is recovered during the purification step.

Fluorescently labeled readout probes have sequences complementary to the readout sequences described above and a Cy5 dye attached at the 3' end. These probes were obtained from IDT and HPLC purified.

Sample preparation and labeling with encoding probes. Human primary fibroblasts (American Type Culture Collection, IMR90) were used in this work. These cells are relatively large and flat, facilitating wide-field imaging without the need for optical sectioning. Cells were cultured with Eagle's Minimum Essential Medium. Cells were plated on 22-mm, #1.5 coverslips (Bioptechs, 0420-0323-2) at 350,000 cells/coverslip and incubated at 37 °C with 5% CO_2 for 48-96 hours within petri dishes. Cells were fixed for 20 minutes in 4% paraformaldehyde (Electron Microscopy Sciences, 15714) in 1x phosphate buffered saline (PBS; Ambion, AM9625) at room temperature, reduced for 5 minutes with 0.1% w/v sodium borohydride (Sigma, 480886) in water to reduce background fluorescence, washed three times with ice-cold 1x PBS, permeabilized for 2 minutes with 0.5% v/v Triton (Sigma, T8787) in 1x PBS at room temperature, and washed three times with ice cold 1x PBS.

Cells were incubated for 5 minutes in encoding wash buffer comprising 2x saline-sodium citrate buffer (SSC) (Ambion, AM9763), 30% v/v formamide (Ambion, AM9342), and 2 mM vanadyl ribonucleoside complex (NEB, S1402S). 10 microliters of 100 micromolar(140-gene experiments) or 200 micromolar (1001-gene experiments) encoding probes in encoding hybridization buffer was added to the cell-containing coverslip and spread uniformly by placing another coverslip on top of the sample. Samples were then incubated in a humid chamber inside a 37 °C-hybridization oven for 18-36 hours. Encoding hybridization buffer is composed of encoding wash buffer supplemented with 1 mg/mL yeast tRNA (Life Technologies, 15401-011) and 10% w/v dextran sulfate (Sigma, D8906-50G).

Cells were then washed with primary encoding wash buffer, incubated at 47 °C for 10 minutes, and this wash was repeated for a total of three times. A 1:1000 dilution of 0.2-micrometer-diameter carboxylate-modified orange fluorescent beads (Life Technologies, F-8809) in 2xSSC was sonicated for 3 minutes and then incubated with the sample for 5

minutes. The beads were used as fiducial markers to align images obtained from multiple successive rounds of hybridization, as described below. The sample was washed once with 2xSSC, and then post-fixed with 4% v/v paraformaldehyde in 2xSSC at room temperature for 30 minutes. The sample was then washed three times with 2xSSC and either imaged immediately or stored for no longer than 12 hours at 4 °C prior to imaging. All solutions were prepared as RNase-free.

MERFISH imaging with multiple successive rounds of hybridization. The sample coverslip was assembled into a Bioptech's FCS2 flow chamber, and the flow through this chamber was controlled via a home-built fluidics system composed of three computer-controlled 8-way valves (Hamilton, MVP and HVXM 8-5) and a computer-controlled peristaltic pump (Rainin, Dynamax RP-1). The sample was imaged on a home-built microscope constructed around an Olympus IX-71 body and a 1.45 NA, 100x oil immersion objective and configured for oblique incidence excitation. The objective was heated to 37 °C with a Bioptechs objective heater. Constant focus was maintained throughout the imaging process with a home-built, auto-focusing system. Illumination was provided at 641 nm, 561 nm, and 405 nm using solid state lasers (MPB communications, VFL-P500-642; Coherent, 561-200CWCDRH; and Coherent, 1069413/AT) for excitation of the Cy5-labeled readout probes, the fiducial beads, and nuclear counterstains, respectively. These lines were combined with a custom dichroic (Chroma, zy405/488/561/647/752RP-UF1) and the emission was filtered with a custom dichroic (Chroma, ZET405/488/561/647-656/752m). Fluorescence was separated with a QuadView (Photometrics) using the dichroics T560lpxr, T650lpxr, 750dcxxr (Chroma) and the emission filters ET525/50m, WT59550m-2f, ET700/75m, HQ770lp (Chroma) and imaged with an EMCCD camera (Andor, iXon-897). The camera was configured so that a pixel corresponds to 167 nm in the sample plane. The entire system was fully automated, so that imaging and fluid handling were performed for the entire experiment without user intervention.

Sequential hybridization, imaging, and bleaching proceeded as follows. 1 mL of 10 nM of the appropriate fluorescently labeled readout probe in readout hybridization buffer (2xSSC; 10% v/v formamide; 10% w/v dextran sulfate, and 2 mM vanadyl ribonucleoside complex) was flown across the sample, flow was stopped, and the sample was incubated for 15 minutes. Then 2 mL of readout wash buffer (2xSSC, 20% v/v formamide; and 2 mM vanadyl ribonucleoside complex) was flown across the sample, flow was stopped, and the sample was incubated for 3 minutes. 2 mL of imaging buffer comprising 2xSSC, 50 mM TrisHCl pH 8, 10% w/v glucose, 2 mM Trolox (Sigma-Aldrich, 238813), 0.5 mg/mL glucose

oxidase (Sigma-Aldrich, G2133), and 40 microgram/mL catalase (Sigma-Aldrich, C30) was flown across the sample. Flow was then stopped, and then approximately 75 to 100 regions were exposed to ~25 mW 642-nm and 1 mW of 561-nm light and imaged. Each region was 40 micrometers by 40 micrometers. The laser powers were measured at the microscope backport. Because the imaging buffer is sensitive to oxygen, the ~50 mL of imaging buffer used for a single experiment was made fresh at the beginning of the experiment and then stored under a layer of mineral oil throughout the measurement. Buffer stored in this fashion was stable for more than 24 hours.

After imaging, the fluorescence of the readout probes was extinguished via photobleaching. The sample was washed with 2 mL of photobleaching buffer (2xSSC and 2 mM vanadyl ribonucleoside complex), and each imaged region of the sample was exposed to 200 mW of 641-nm light for 3 s. To confirm the efficacy of this photobleaching treatment, imaging buffer was reintroduced, and the sample was imaged as described above.

The above hybridization, imaging, and photobleaching process was repeated either 16 times for the 140-gene measurements using the MHD4 code or 14 times for the 1001-gene measurements using the MHD2 code. An entire experiment was typically completed in ~20 hours.

Following completion of imaging, 2 mL of a 1:1000 dilution of Hoescht (ENZ-52401) in 2xSSC was flown through the chamber to label the nuclei of the cells. The sample was then washed immediately with 2 mL of 2xSSC followed by 2 mL of imaging buffer. Each region of the sample was then imaged once again with ~1 mW of 405-nm light.

Because cells were imaged using wide-field imaging with oblique-incidence illumination, without optical sectioning and z-scanning, the fraction of individual RNA species that was outside the axial range of the imaging geometry was quantified for 6 different RNA species using conventional smFISH. For this purpose, these cells were optically sectioned by collecting stacks of images at different focal depths through the entire depth of the cells. The images were aligned in consecutive focal planes and then computed for each cell the fraction of RNAs that were detected in the three-dimensional stack but not in the basal focal plane. It was found that only a small fraction, 15% +/- 1% (Mean +/- SEM across six different RNA species) of RNA molecules were outside the imaging range of a fixed focal plane without z-scanning. These measurements also confirmed that the excitation geometry illuminated the full depth of the cells. Any optical sectioning technique could be employed in MERFISH to allow the imaging of RNAs in thicker cells or tissues.

Construction of measured words. Fluorescent spots were identified and localized in each image using a multi-Gaussian-fitting algorithm assuming a Gaussian with a uniform width of 167 nm. This algorithm was used to allow partially overlapping spots to be distinguished and individually fit. RNA spots were distinguished from background signal, i.e. signal arising from probes bound non-specifically, by setting the intensity threshold required to fit a spot with this software. Due to variation in the brightness of spots between rounds of hybridization, this threshold was adjusted appropriately for each hybridization round to minimize the combined average of the 1-->0 and 0-->1 error rates across all hybridization rounds (140-gene measurements) or to maximize the ratio of the number of measured words with four '1' bits to those with three or five '1' bits (1001-gene measurements). The location of the fiducial beads was identified in each frame using a faster single-Gaussian fitting algorithm.

Images of the same sample region in different rounds of hybridization were registered by rotating and translating the image to align the two fiducial beads within the same image that were most similar in location after a coarse initial alignment via image correlation. All images were aligned to a coordinate system established by the images collected in the first round of hybridization. The quality of this alignment was determined from the residual distance between five additional fiducial beads, and alignment error was typically ~20 nm.

Fluorescence spots in different hybridization rounds were connected into a single string, corresponding to a potential RNA molecule, if the distance between spots was smaller than 1 pixel (167 nm). For each string of spots, the on-off sequence of fluorescent signals in all hybridization rounds were used to assign a binary word to the potential RNA molecule, in which '1' was assigned to the hybridization rounds that contained a fluorescent signal above threshold and '0' was assigned to the other hybridization rounds. Measured words were then decoded into RNA species using the 16-bit MHD4 code or the 14-bit MHD2 code discussed above. In the case of the 16-bit MHD4 code, if the measured binary word matched the code word of a specific RNA perfectly or differed from the code word by one single bit, it was assigned to that RNA. In the case of the 14-bit MHD2 code, only if the measured binary word matched the code word of a specific RNA perfectly, was it assigned to that RNA. To determine the copy number per cell, the number of each RNA species was counted in individual cells within each 40 micrometer by 40 micrometer imaging area. It is noted that this number accounts for the majority but not all RNA molecules within a cell because a fraction of the cell could be outside the imaging area or focal depth. Tiling images of

adjacent areas and adjacent focal planes could be employed to improve the counting accuracy.

In the 140-gene experiments, some regions of the cell nucleus occasionally contained too much fluorescence signal to properly identify individual RNA spots. In the 1001-gene experiments, the cell nucleus generally contained too much fluorescent signal to allow identification of individual RNA molecules. These bright regions were excluded from all subsequent analysis. This work focuses on mRNAs, which are enriched in the cytoplasm. To estimate the fraction of mRNAs missed by excluding the nucleus region, conventional smFISH was used to quantify the fraction of molecules found inside the nucleus for six different mRNAs species. It was found that only 5% +/- 2% (Mean +/- SEM across six RNA species) of these RNA molecules are found in the nucleus. Employment of super-resolution imaging and/or optical sectioning could potentially allow individual molecules in these dense nucleus regions to be identified, which would be particularly useful for probing those non-coding RNAs that are enriched in the nucleus.

smFISH measurements of individual genes. Pools of 48 fluorescently-labeled (Quasar 670) oligonucleotide probes per RNA were purchased from Biosearch Technologies. 30-nt probe sequences were taken directly from a random subset of the targeting regions used for the multiplexed measurements. Cells were fixed and permeabilized as described above. 10 microliters of 250 nM oligonucleotide probes in encoding hybridization buffer (described above) was added to the cell-containing coverslip and spread uniformly by placing another coverslip on top of the sample. Samples were then incubated in a humid chamber inside a 37 ^{o}C-hybridization oven for 18 hours. Cells were then washed with encoding wash buffer (described above) at 37 ^{o}C for 10 minutes, and this wash was repeated for a total of three times. The sample was then washed three times with 2xSSC and imaged in imaging buffer using the same imaging geometry as described above for MERFISH.

Bulk RNA sequencing. Total RNA was extracted from IMR90 cells cultured as above using the Zymo Quick RNA MiniPrep kit (R1054) according to the manufacturer's instructions. polyA RNA was then selected (NEB; E7490), and a sequencing library was constructed using the NEBNext Ultra RNA library preparation kit (NEB; E7530), amplified with custom oligonucleotides, and 150-bp reads were obtained from on a MiSeq. These sequences were aligned to the human genome (Gencode v18) and isoform abundance was computed with cufflinks.

Calculation of the predicted scaling and error properties of different encoding schemes. Analytic expressions were derived for the dependence of the number of possible

code words, the calling rate, and the misidentification rate on *N*. The calling rate is defined as the fraction of RNA molecules that are properly identified. The misidentification rate is defined as the fraction of RNA molecules that are misidentified as a wrong RNA species. For encoding schemes with an error-detection capability, the calling rate and misidentification rate does not add up to 1 because a fraction of the molecules not called properly can be detected as errors and discarded and, hence, not misidentified as a wrong species. These calculations assume that the probability of misreading bits is constant for all hybridization rounds but differs for the 1-->0 and 0-->1 errors. Experimentally measured average 1-->0 and 0-->1 error rates (10% and 4% respectively) were used for the estimates shown in Figs. 5B-5D. For simplicity, the word corresponding to all '0's was not removed from calculations.

For the simple binary encoding scheme in which all possible *N*-bit binary words are assigned to unique RNA species, the number of possible code words is 2^N. The number of words that could be used to encode RNA is actually $2^N - 1$ because the code word '00...0' does not contain detectable fluorescence in any hybridization round, but for simplicity the word corresponding to all '0's was not removed from subsequent calculations. The error introduced by this approximation is negligible. For any given word with *m* '1's and *N*–*m* '0's the probability of measuring that word without error (the fraction of RNAs that is properly called) is:

$$(1-p_1)^m(1-p_0)^{N-m}, \tag{1}$$

where p_1 is 1-->0 error rate and p_0 is 0-->1 error rate per bit. Because different words in this simple binary encoding scheme can have different numbers of '1' bits, the calling rate for different words will differ if $p_1 \neq p_0$. The average calling rate, reported in Fig. 5C, was determined from the weighted average of the value of Eq. (1) for all words. This weighted average is:

$$\frac{1}{2^N}\sum_{m=0}^{N}\binom{N}{m}(1-p_1)^m(1-p_0)^{N-m}, \tag{2}$$

where $\binom{N}{m}$ is the binomial coefficient and corresponds to the number of words with *m* '1' bits in this encoding scheme. Since in this encoding scheme every error produces a binary word that encodes a different RNA, the average misidentification rate for this encoding scheme, reported in Fig. 5D, follows directly from (2):

$$1-\frac{1}{2^N}\sum_{m=0}^{N}\binom{N}{m}(1-p_1)^m(1-p_0)^{N-m}. \quad (3)$$

To calculate the scaling and error properties of the extended Hamming distance 4 (HD4) code, the generator matrix for the desired number of data bits using standard methods was first created. The generator matrix determines the specific words that are present in a given encoding scheme and was used to directly determine the number of encoded words as a function of the number of bits. In this encoding scheme, the calling rate corresponds to the fraction of words measured without error as well as the fraction of words measured with a single-bit error. For code words with *m* '1' bits, this fraction is determined by the following expression:

$$(1-p_1)^m(1-p_0)^{N-m}+m\,p_1{}^1(1-p_1)^{m-1}(1-p_0)^{N-m}+(N-m)\,p_0{}^1(1-p_1)^m(1-p_0)^{N-m-1} \quad (4)$$

where the first term is the probability of not making any errors, the second term corresponds to the total probability of making one 1-->0 error at any of the *m* '1' bits without making any other 0-->1 errors, and the final term corresponds to the total probability of making one 0-->1 error at any of the *N-m* '0' bits without making any 1-->0 errors. Because the number of '1' bits can differ between words in this encoding scheme, the average calling rate reported in Fig. 5C was computed from a weighted average over Eq. (4) for different values of *m*. The weight for each term was determined from the number of words that contain *m* '1' bits as determined from the generator matrix described above.

Because RNA-encoding words are separated by a minimum Hamming distance of 4, at least 4 errors are required to switch one word into another. If error correction is applied, then 3 or 5 errors could also convert one RNA into another. Thus, the misidentification rate from all possible combinations of 3-bit, 4-bit and 5-bit errors was estimated for code words with *m* '1' bits. Technically, >5-bit errors could also convert one RNA into another, but the probability of making such errors is negligible because of the small per-bit error rate. This expression was approximated with:

$$\begin{aligned}
&\sum_{i=0}^{4}\binom{m}{i}\binom{N-m}{4-i}p_1^i p_0^{4-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(4-i)}+\\
&\sum_{i=0}^{3}\binom{m}{i}\binom{N-m}{3-i}p_1^i p_0^{3-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(3-i)}+\\
&\sum_{i=0}^{5}\binom{m}{i}\binom{N-m}{5-i}p_1^i p_0^{5-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(5-i)}+
\end{aligned} \quad (5)$$

The first sum corresponds to all of the ways in which exactly four mistakes can be made. Similarly, the second and third sums correspond to all of the ways in which exactly three or

five mistakes can be made. Eq. (5) provides an upper bound for the misidentification rate because not all three, four, or five bit errors produce a word that matches or would be corrected to another legitimate word. Again because the number of '1' bits can differ between words, the average misidentification rate reported in Fig. 5D is calculated as a weighted average of Eq. (5) over the number of words that have *m* '1' bits.

To generate the MHD4 code where the number of '1' bits for each code word is set to 4, the HD4 codes were first generated as described above, and then all code words that did not contain four '1's were removed. The calling rate of this code, reported in Fig. 5C, was directly calculated from Eq. (4) but with m = 4 because all code words in this code have four '1' bits. The misidentification rate of this code, reported in Fig. 5D, was calculated by modifying Eq. (5) with the following considerations: (i) the number of '1' bits, m, was set to 4 and (ii) errors that produce words that do not contain three, four, or five '1' bits were excluded. Thus, the expression in Eq. (5) was simplified to

$$\begin{aligned}
&\binom{4}{2}\binom{N-4}{2}p_1^2 p_0^2(1-p_1)^2(1-p_0)^{N-6} + \\
&\binom{4}{1}\binom{N-4}{2}p_1^1 p_0^2(1-p_1)^3(1-p_0)^{N-6} + \binom{4}{2}\binom{N-4}{1}p_1^2 p_0^1(1-p_1)^2(1-p_0)^{N-5} \\
&\binom{4}{2}\binom{N-4}{3}p_1^2 p_0^3(1-p_1)^2(1-p_0)^{N-7} + \binom{4}{3}\binom{N-4}{2}p_1^3 p_0^2(1-p_1)^1(1-p_0)^{N-6}
\end{aligned} \tag{6}$$

Again, this expression is an upper bound on the actual misidentification rate because not all words with four '1's are valid code words.

Estimates of the 1-->0 and 0-->1 error rates for each hybridization round. To compute the probability of misreading a bit at a given hybridization round, the error correcting properties of the MHD4 code were used. Briefly, the probabilities of 1-->0 or 0-->1 errors were derived in the following way. Let the probability of making an error at the *i*th bit, i.e. *i*th hybridization round, be p_i and the actual number of RNA molecules of the given species be *A*, then the number of exact matches for this RNA will be $W_E = A\prod_{i=1}^{16}(1-p_i)$ and the number of one-bit error corrected matches for this RNA corresponding to errors at the *i*th bit will be $W_i = A\frac{p_i}{(1-p_i)}\prod_{j=1}^{16}(1-p_j)$. The p_i can be directly derived from the ratio: $W_i/W_E = \frac{p_i}{(1-p_i)}$. This ratio assumes that the one-bit error-corrected counts were only generated from single-bit errors from the correct word and that multi-error contamination from other RNA words is negligible. Given that the error rate per hybridization round is small and that it takes at least

three errors to convert one RNA-encoding word into a word that would be misidentified as another RNA, the above approximation should be a good one.

To compute the average 1-->0 or 0-->1 error probabilities for each of the 16 hybridization rounds, the above approach was used to calculate the per-bit error rates for each bit of every gene, and these errors were sorted based on whether they correspond to a 1-->0 or a 0-->1 error, and the average of these errors for each bit weighted by the number of counts observed for the corresponding gene was taken.

Estimates of the calling rate for individual RNA species from actual imaging data. With the estimates of the 1-->0 or 0-->1 error probabilities for each round of hybridization as determined above, it is possible to estimate the calling rate for each RNA based on the specific word used to encoded it. Specifically, the fraction of an RNA species that is called correctly is determined by

$$\prod_{i=1}^{N}(1-p_i)+\sum_{j=1}^{N}\frac{p_j}{(1-p_j)}\prod_{i=1}^{N}(1-p_i), \qquad (7)$$

where the first term represent the probability of observing an exact match of the code word and the second term represent the probability of observing an error-corrected match (i.e. with one-bit error). The values of the per-bit error rate p_i for each RNA species are determined by the specific code word for that RNA and the measured 1-->0 or 0-->1 error rates for each round of hybridization. If the code word of the RNA contains a '1' in the *i*th bit, then p_i is determined from the 1-->0 error rate for the *i*th hybridization round; if the word contains a '0' in the *i*th bit, p_i is determined from the 0-->1 error rate for the *i*th hybridization round.

Hierarchical clustering analysis of the co-variation in RNA abundance. Hierarchical clustering of the co-variation in gene expression for both the 140-gene and 1001-gene experiments was conducted as follows. First, the distance between every pair of genes was determined as 1 minus the Pearson correlation coefficient of the cell-to-cell variation of the measured copy numbers of these two RNA species, both normalized by the total RNA counted in the cell. Thus, highly correlated genes are 'closer' to one another and highly anti-correlated genes are 'further' apart. An agglomerative hierarchical cluster tree was then constructed from these distances using the Unweighted Pair Group Method with Arithmetic mean (UPGMA). Specifically, starting with individual genes, hierarchical clusters were constructed by identifying the two clusters (or individual genes) that are closest to one another according to the arithmetic mean of the distances between all inter-cluster gene pairs. The pairs of clusters (or individual genes) with the smallest distance are then grouped

together and the process is repeated. The matrix of pairwise correlations was then sorted based on the order of the genes within these trees.

Groups of genes with substantial co-variations were identified by selecting a threshold on the hierarchical cluster tree (indicated by the dashed lines in Figs. 7D and 10A) that produced approximately 10 groups of genes each of which contains at least 4 members for the 140-gene experiments or approximately 100 groups each of which contains at least 3 members for the 1001-gene experiments. It is noted that one can change the threshold in order to identify either more tightly coupled smaller groups or larger groups with relatively loose coupling.

A probability value for the confidence that a gene belongs to a specific group was determined by computing the difference between the average correlation coefficient between that gene and all other members of that group and the average correlation coefficient between that gene and all other measured genes outside that group. The significance (p-value) of this difference was determined with the student's t-test.

Because hierarchical clustering is inherently a one-dimensional analysis, i.e. any given genes can only be a member of a single group, this analysis does not allow all correlated gene groups to be identified. Higher dimension analysis, such as principal component analysis or k-means clustering, could be used to identify more co-varying gene clusters.

Analysis of RNA spatial distributions. To identify genes that have similar spatial distributions, each of the measured cells was subdivided into 2x2 regions and calculated the fraction of each RNA species present in each of these bins. To control for the fact that some regions of the cell naturally contain more RNA than others, the enrichment was calculated for each gene , i.e., the ratio of the observed fraction in a given region for a given RNA species to the average fraction observed for all genes in that same region. For each pair of RNA species, the Pearson correlation coefficient of the region-to-region variation in enrichment of these two RNA species for each cell was determined and the correlation coefficients were averaged over ~400 cells imaged in 7 independent data sets. RNA species were then clustered based on these average correlation coefficients using the same hierarchical clustering algorithm described above. Because of the large number of cells used for the analysis, it was found that the coarse spatial binning (2x2 regions per cell) was sufficient to capture the spatial correlation between genes and finer binning did not produce more significantly correlated groups.

To measure the distances of genes from the nuclei and from the cell edge, brightness thresholds on the cell images were first used to segment the nuclei and the cell edges identified. The distance from every RNA molecule to the nearest part of the nucleus and nearest part of the cell edge was then determined. For each data set, the average distance for each RNA species averaged over all the cells measured was then computed. These distances were averaged for the group I genes, group II genes or all genes. Only those RNA species with at least 10 counts per cell were used in this analysis to minimize statistical error on the distance values.

Gene ontology (GO) analysis. Groups of genes were selected from the hierarchical trees as discussed above. A collection of GO terms was determined for all measured RNA species as well as the RNA species associated with each group from the most recent human GO annotations using both the annotated GO terms and terms immediately upstream or downstream of the found annotations. The enrichment of these annotations was calculated from the ratio of the fraction of genes within each group that have this term to the fraction of all measured genes that have this term and the p-value for this enrichment was calculated via the hypergeometric function. Only statistically significantly enriched GO terms with a p-value less than 0.05 were considered.

While several embodiments of the present invention have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the functions and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the present invention. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the teachings of the present invention is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, the invention may be practiced otherwise than as specifically described and claimed. The present invention is directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles,

materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the scope of the present invention.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

The indefinite articles "a" and "an," as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean "at least one."

The phrase "and/or," as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with "and/or" should be construed in the same fashion, i.e., "one or more" of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the "and/or" clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to "A and/or B", when used in conjunction with open-ended language such as "comprising" can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, "or" should be understood to have the same meaning as "and/or" as defined above. For example, when separating items in a list, "or" or "and/or" shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as "only one of" or "exactly one of," or, when used in the claims, "consisting of," will refer to the inclusion of exactly one element of a number or list of elements. In general, the term "or" as used herein shall only be interpreted as indicating exclusive alternatives (i.e. "one or the other but not both") when preceded by terms of exclusivity, such as "either," "one of," "only one of," or "exactly one of." "Consisting essentially of," when used in the claims, shall have its ordinary meaning as used in the field of patent law.

As used herein in the specification and in the claims, the phrase "at least one," in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also

allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase "at least one" refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, "at least one of A and B" (or, equivalently, "at least one of A or B," or, equivalently "at least one of A and/or B") can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

When the word "about" is used herein in reference to a number, it should be understood that still another embodiment of the invention includes that number not modified by the presence of the word "about."

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

In the claims, as well as in the specification above, all transitional phrases such as "comprising," "including," "carrying," "having," "containing," "involving," "holding," "composed of," and the like are to be understood to be open-ended, i.e., to mean including but not limited to. Only the transitional phrases "consisting of" and "consisting essentially of" shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

RELATED APPLICATIONS

This application is a national stage filing of International Patent Application Serial No. PCT/US2015/042556, filed July 29, 2017, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*, which claims the benefit of U.S. Provisional Patent Application Serial No. 62/031,062, filed July 30, 2014, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/142,653, filed April 3, 2015, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; and U.S. Provisional Patent Application Serial No. 62/050,636, filed September 15, 2014, entitled "Probe Library Construction," by Zhuang, *et al.* Each of the above is incorporated herein by reference.

GOVERNMENT FUNDING

This invention was made with government support under Grant No. GM096450 awarded by National Institutes of Health. The government has certain rights in the invention.

FIELD

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined.

BACKGROUND

Single-molecule fluorescent in situ hybridization (smFISH) is a powerful method for detecting individual mRNA molecules in cells. The high detection efficiency and large dynamic range of this method provides exquisite detail into the expression state, spatial distribution within cells and intact tissues, and variation among cells of individual mRNAs. Such approaches have been essential to many recent insights into understanding gene regulation and expression. A fundamental limitation of smFISH, however, is its low throughput, typically only a few genes at a time. This low throughput is due to a lack of distinguishable probes with which to label cells and the cost of producing large amounts of labeled probe required for high efficiency staining. Thus, improvements in detecting mRNA molecules are needed.

SUMMARY

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. The subject matter of the present invention

involves, in some cases, interrelated products, alternative solutions to a particular problem, and/or a plurality of different uses of one or more systems and/or articles.

In one aspect, the present invention is generally directed to a composition. According to one set of embodiments, the composition comprises a plurality of nucleic acid probes, at least some of which comprise a first portion comprising a target sequence and a plurality of read sequences. In some cases, each comprises a first portion comprising a target sequence and a plurality of read sequences. In some embodiments, the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

In another aspect, the present invention is generally directed to a method. In one set of embodiments, the method includes acts of exposing a sample to a plurality of nucleic acid probes; for each of the nucleic acid probes, determining binding of the nucleic acid probes within the sample; creating codewords based on the binding of the nucleic acid probes; and for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

The method, in another set of embodiments, includes acts of exposing a sample to a plurality of nucleic acid probes, wherein the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences, and wherein at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences; and for each of the nucleic acid probes, determining binding of the target sequences of the nucleic acid probes within the sample.

In yet another set of embodiments, the method includes acts of exposing a sample to a plurality of primary nucleic acid probes (also called encoding probes); exposing the plurality of primary nucleic acid probes to a sequence of secondary nucleic acid probes (also called readout probes) and determining fluorescence of each of the secondary nucleic acid probes within the sample; creating codewords based on fluorescence of the secondary nucleic acid probes; and for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

In one set of embodiments, the method includes acts of exposing a plurality of primary nucleic acid probes to a sample; and exposing the plurality of nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary probes within the sample. In some embodiments, at least some of the plurality of secondary nucleic acid probes comprises distinguishable secondary nucleic acid probes

formed from combinatorial combination of one or more read sequences (or readout probe sequences) taken from a plurality of read sequences (or readout probe sequences).

In another set of embodiments, the method comprises acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a precision better than 500 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a precision better than 500 nm.

In another set of embodiments, the method comprises acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a resolution better than 100 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a resolution better than 100 nm.

The method, in yet another set of embodiments, includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity using a super-resolution imaging technique, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In certain embodiments, the method comprises acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords; and forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

In addition, in one set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes each comprising a target sequence; and forming groups comprising the

plurality of nucleic acid probes such that each group of nucleic acid probes corresponds to at least one common value of a position within the codewords.

In another set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions that is less than the number of targets, and wherein each codeword is associated with a single target, associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords, and forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

The method, in still another set of embodiments, includes acts of exposing a plurality of nucleic acid probes to a cell, exposing the plurality of nucleic acid probes to a sequence of secondary probes and determining fluorescence of each of the secondary probes within the cell, and based on the sequence of fluorescence of each of the secondary probes, determining nucleic acids within the cell.

In another set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes each comprising a target sequence; and forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes correspond to at least one common value of a position within the codewords.

In yet another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity using a super-resolution imaging technique, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In still another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a precision better than 500 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a precision better than 500 nm.

In still another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a resolution better than 100 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In still another set of embodiments, the method includes acts of associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code; associating unique read sequences with each possible value of each position in the codewords, wherein the read sequences are taken from a set of orthogonal sequences, which have limited homology with one another and with the nucleic acid species in a sample; forming a plurality of primary nucleic acid probes each comprising a target sequence that uniquely binds to a nucleic acid target and one or more read sequences; forming a plurality of secondary nucleic acid probes comprising a signaling entity and a sequence that is complementary to one of the read sequences; exposing a sample to the primary nucleic acid probes such that the nucleic acid probes hybridize to the nucleic acid targets in the sample; exposing the primary nucleic acid probes in the sample to a secondary nucleic acid probe such that the secondary nucleic acid probe hybridizes to the read sequence on at least some of the primary nucleic acid probes; imaging the sample; and repeating the exposing and imaging steps one or more times, using a different secondary nucleic acid probe for at least some of the repetitions.

The method, according to yet another set of embodiments, includes acts of associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes comprising a signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets; grouping the nucleic acid probes into a plurality of probe pools, wherein each of the probe pools corresponds to a specific value of a unique position within the codewords; exposing a sample to one of the probe pools; imaging the sample; and repeating the exposing and imaging steps one or more times, using a different probe pool for at least some of the repetitions.

In another aspect, the present invention encompasses methods of making one or more of the embodiments described herein. In still another aspect, the present invention encompasses methods of using one or more of the embodiments described herein.

Other advantages and novel features of the present invention will become apparent from the following detailed description of various non-limiting embodiments of the invention when considered in conjunction with the accompanying figures. In cases where the present specification and a document incorporated by reference include conflicting and/or inconsistent disclosure, the present specification shall control. If two or more documents incorporated by reference include conflicting and/or inconsistent disclosure with respect to each other, then the document having the later effective date shall control.

BRIEF DESCRIPTION OF THE DRAWINGS

Non-limiting embodiments of the present invention will be described by way of example with reference to the accompanying figures, which are schematic and are not intended to be drawn to scale. In the figures, each identical or nearly identical component illustrated is typically represented by a single numeral. For purposes of clarity, not every component is labeled in every figure, nor is every component of each embodiment of the invention shown where illustration is not necessary to allow those of ordinary skill in the art to understand the invention. In the figures:

Figs. 1A-1C illustrates an encoding scheme for nucleic acid probes, in certain embodiments of the invention;

Figs. 2A-2G illustrates determining of mRNAs in a cell, in some embodiments of the invention;

Figs. 3A-3B illustrate the determination of nucleic acids, in accordance with various embodiments of the invention;

Figs. 4A-4B is a non-limiting example of multiple read sequences distributed in a population of different nucleic acid probes, in accordance with certain embodiments of the invention;

Figs. 5A-5E illustrate the determination of nucleic acids, in accordance with another embodiment of the invention;

Figs. 6A-6H illustrate simultaneous determination of multiple nucleic acid species in cells, in certain embodiments of the invention;

Figs. 7A-7E show expression noise of genes and co-variation of expression between different genes determined in accordance with some embodiments of the invention;

Figs. 8A-8E illustrate spatial distribution of RNAs in cells determined in accordance with one embodiment of the invention;

Figs. 9A-9C illustrate simultaneous determination of multiple nucleic acid species in cells, in another embodiment of the invention;

Figs. 10A-10B show expression between different genes determined in accordance with yet another embodiment of the invention;

Fig. 11 is a schematic description of combinatorial labeling, in accordance with another embodiment of the invention;

Figs. 12A-12C show schematic descriptions of Hamming distance, in another embodiment of the invention;

Fig. 13 illustrates the production of a library of probes, in still another embodiment of the invention;

Figs. 14A-14B illustrate fluorescent spot determinations, in another embodiment of the invention;

Figs. 15A-15B illustrate error correction facilitates RNA detection, in yet another embodiment of the invention;

Figs. 16A-16E show characterization of misidentification rates and calling rates, in one embodiment of the invention;

Figs. 17A-17D show characterization of misidentification rates and calling rates, in another embodiment of the invention;

Figs. 18A-18C show a comparison of experiments, in accordance with one embodiment of the invention;

Figs. 19A-19D illustrate decoding and error assessment, in another embodiment of the invention; and

Figs. 20A to 20H illustrate encoding of 130 RNA species and 10 code words as misidentification controls as Codebook 1 and Codebook 2, in another embodiment of the invention.

BRIEF DESCRIPTION OF THE SEQUENCES

SEQ ID NO: 1 is:

GTTGGCGACGAAAGCACTGCGATTGGAACCGTCCCAAGCGTTGCGCTTAATGGA TCATCAATTTTGTCTCACTACGACGGTCAATCGCGCTGCATACTTGCGTCGGTCG GACAAACGAGG;

SEQ ID NO: 2 is CGCAACGCTTGGGACGGTTCCAATCGGATC;

SEQ ID NO: 3 is CGAATGCTCTGGCCTCGAACGAACGATAGC;

SEQ ID NO: 4 is ACAAATCCGACCAGATCGGACGATCATGGG;

SEQ ID NO: 5 is CAAGTATGCAGCGCGATTGACCGTCTCGTT;

SEQ ID NO: 6 is TGCGTCGTCTGGCTAGCACGGCACGCAAAT;

SEQ ID NO: 7 is AAGTCGTACGCCGATGCGCAGCAATTCACT;

SEQ ID NO: 8 is CGAAACATCGGCCACGGTCCCGTTGAACTT;

SEQ ID NO: 9 is ACGAATCCACCGTCCAGCGCGTCAAACAGA;

SEQ ID NO: 10 is CGCGAAATCCCCGTAACGAGCGTCCCTTGC;

SEQ ID NO: 11 is GCATGAGTTGCCTGGCGTTGCGACGACTAA;

SEQ ID NO: 12 is CCGTCGTCTCCGGTCCACCGTTGCGCTTAC;

SEQ ID NO: 13 is GGCCAATGGCCCAGGTCCGTCACGCAATTT;

SEQ ID NO: 14 is TTGATCGAATCGGAGCGTAGCGGAATCTGC;

SEQ ID NO: 15 is CGCGCGGATCCGCTTGTCGGGAACGGATAC;

SEQ ID NO: 16 is GCCTCGATTACGACGGATGTAATTCGGCCG;

SEQ ID NO: 17 is GCCCGTATTCCCGCTTGCGAGTAGGGCAAT

SEQ ID NO: 18 is GTTGGTCGGCACTTGGGTGC;

SEQ ID NO: 19 is CGATGCGCCAATTCCGGTTC;

SEQ ID NO: 20 is CGCGGGCTATATGCGAACCG;

SEQ ID NO: 21 is
TAATACGACTCACTATAGGGAAAGCCGGTTCATCCGGTGG;

SEQ ID NO: 22 is
TAATACGACTCACTATAGGGTGATCATCGCTCGCGGGTTG;

SEQ ID NO: 23 is
TAATACGACTCACTATAGGGCGTGGAGGGCATACAACGC;

SEQ ID NO: 24 is CGCAACGCTTGGGACGGTTCCAATCGGATC/3Cy5Sp/;

SEQ ID NO: 25 is CGAATGCTCTGGCCTCGAACGAACGATAGC/3Cy5Sp/;

SEQ ID NO: 26 is ACAAATCCGACCAGATCGGACGATCATGGG/3Cy5Sp/;

SEQ ID NO: 27 is CAAGTATGCAGCGCGATTGACCGTCTCGTT/3Cy5Sp/;

SEQ ID NO: 28 is GCGGGAAGCACGTGGATTAGGGCATCGACC/3Cy5Sp/;

SEQ ID NO: 29 is AAGTCGTACGCCGATGCGCAGCAATTCACT/3Cy5Sp/;

SEQ ID NO: 30 is CGAAACATCGGCCACGGTCCCGTTGAACTT/3Cy5Sp/;

SEQ ID NO: 31 is ACGAATCCACCGTCCAGCGCGTCAAACAGA/3Cy5Sp/;

SEQ ID NO: 32 is CGCGAAATCCCCGTAACGAGCGTCCCTTGC/3Cy5Sp/;

SEQ ID NO: 33 is GCATGAGTTGCCTGGCGTTGCGACGACTAA/3Cy5Sp/;

SEQ ID NO: 34 is CCGTCGTCTCCGGTCCACCGTTGCGCTTAC/3Cy5Sp/;

SEQ ID NO: 35 is GGCCAATGGCCCAGGTCCGTCACGCAATTT/3Cy5Sp/;
SEQ ID NO: 36 is TTGATCGAATCGGAGCGTAGCGGAATCTGC/3Cy5Sp/;
SEQ ID NO: 37 is CGCGCGGATCCGCTTGTCGGGAACGGATAC/3Cy5Sp/;
SEQ ID NO: 38 is GCCTCGATTACGACGGATGTAATTCGGCCG/3Cy5Sp/; and
SEQ ID NO: 39 is GCCCGTATTCCCGCTTGCGAGTAGGGCAAT/3Cy5Sp/

.

DETAILED DESCRIPTION

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. Certain embodiments are directed to determining nucleic acids, such as mRNA, within cells at relatively high resolutions. In some embodiments, a plurality of nucleic acid probes may be applied to a sample, and their binding within the sample determined, e.g., using fluorescence, to determine locations of the nucleic acid probes within the sample. In some embodiments, codewords may be based on the binding of the plurality of nucleic acid probes, and in some cases, the codewords may define an error-correcting code to reduce or prevent misidentification of the nucleic acids. In certain cases, a relatively large number of different targets may be identified using a relatively small number of labels, e.g., by using various combinatorial approaches.

Two example approaches are now discussed. It should be understood, however, that these are presented by way of explanation and not limitation; other aspects and embodiments are discussed in further detail herein. In one example method, primary probes (also called encoding probes) and secondary probes (also called readout probes) are used, where the primary probes encode "codewords" and bind to target nucleic acids in the sample, and the secondary probes are used to read out the codewords from the primary probes. In another example method, a plurality of different primary probes containing codewords are divided into as many separate pools as there are positions in the codewords, such that each primary probe pool corresponds to a certain value in a certain position of the codewords (e.g., a "one" in the first position as in "1001").

The first example is now described with respect to Fig. 3A. As will be discussed in more detail below, in other embodiments, other configurations may be used as well. In this first example, a series of nucleic acid probes are used to determine nucleic acids within a cell or other sample, e.g., qualitatively or quantitatively. For example, nucleic acids may be identified as being present or absent, and/or the numbers or concentrations of certain nucleic acids may be determined within the cell or other sample. In some cases, the positions of the

probes within the cell or other sample may be determined at relatively high resolutions, and in some cases, at resolutions better than the wavelength of visible light.

This example is generally directed to spatially detecting nucleic acids within a cell or other sample, e.g., at relatively high resolutions. For example, the nucleic acids may be mRNAs, or other nucleic acids described herein. In one set of embodiments, the nucleic acids within the cell may be determined by delivering or applying nucleic acid probes to the cell. In some cases, by using combinatorial approaches, a relatively large number of nucleic acids may be determined using a relatively small number of different labels on the nucleic acid probes. Thus, for example, a relatively small number of experiments may be used to determine a relatively large number of nucleic acids in a sample, e.g., due to simultaneous binding of the nucleic acid probes to different nucleic acids in the sample.

In one set of embodiments, a population of primary nucleic acid probes are applied to the cell (or other sample) that is able to bind nucleic acids suspected of being present within the cell. Afterwards, sequentially, secondary nucleic acid probes that can bind to or otherwise interact with some of the primary nucleic acids are added and determined, e.g., using imaging techniques such as fluorescence microscopy (e.g., conventional fluorescence microscopy), STORM (stochastic optical reconstruction microscopy) or other imaging techniques. After imaging, the secondary nucleic acid probes are inactivated or removed, and a different secondary nucleic acid probe is added to the sample. This may be repeated multiple times with multiple different secondary nucleic acid probes. The pattern of binding of the various secondary nucleic acid probes may be used to determine the primary nucleic acid probes at locations within the cell or other sample, which can be used to determine mRNA or other nucleic acids that are present.

For instance, as is shown in Fig. 3A, a population of nucleic acids 10 within a cell (represented here by nucleic acids 11, 12, and 13) may be exposed to a population of primary nucleic acid probes 20, including probes 21 and 22. The primary nucleic acid probes may contain, for instance, a target sequence that can recognize a nucleic acid (e.g., a sequence within nucleic acid 11). Probes 21 and 22 may contain the same or different targeting sequences, which may bind to or hybridize with the same or different nucleic acids. As an example, as is shown in Fig. 3A, probe 21 contains a first targeting sequence 25 which targets the probe to nucleic acid 11, while probe 22 contains a second targeting sequence 26, not identical to the first targeting sequence 25 and which targets the probe to nucleic acid 12. The target sequence may be substantially complementary to at least a portion of a target

nucleic acid, and enough of the target sequence may be present such that specific binding of the nucleic acid probe to the target nucleic acid can occur.

Primary nucleic acid probes 20 may also contain one or more "read" sequences. Two such read sequences are used in this example, although in other embodiments, there may be one, three, four, or more read sequences present within a primary nucleic acid probe. The read sequences may all independently be the same or different. In addition, in one set of embodiments, different nucleic acid probes may use one or more common read sequences. For example, more than one read sequence may be combinatorially combined on different nucleic acid probes, thereby producing a relatively large number of different nucleic acid probes that can be separately identified, even though only a relatively small number of read sequences are used. Thus, for example, in Fig. 3A, probe 21 contains read sequences 27 and 29, while probe 22 contains read sequences 27 and 28, where the two read sequences 27 are identical, and different from read sequences 28 and 29.

After primary nucleic acid probes 20 have been introduced to the sample and allowed to interact with nucleic acids 11, 12, and 13, one or more secondary nucleic acid probes 30 may be applied to the sample to determine the primary nucleic acid probes. The secondary nucleic acid probes may contain a recognition sequence able to recognize one of the read sequences present within the population of primary nucleic acid probes. For instance, the recognition sequence may be substantially complementary to at least a portion of the read sequence, such that the secondary nucleic acid probe is able to bind to or hybridize with corresponding primary nucleic acid probe. For instance, in this example, recognition sequence 35 is able to recognize read sequence 27. In addition, the secondary nucleic acid probes may contain one or more signaling entities 33. For example, a signaling entity may be a fluorescent entity attached to the probe, or a certain sequence of nucleic acids that can be determined in some fashion. More than one secondary sequence may be used, e.g., sequentially. For example, as shown in this figure, the initial secondary probe 30 may be removed (e.g., as discussed below) and a new secondary probe 31 may be added, containing recognition sequence 36 able to recognize read sequence 28 and one or more signaling entities 33. This may also be repeated multiple times, e.g., to determine read sequence 29 or other read sequences that may be present.

The location of the secondary nucleic acid probes 30, 31, etc. may be determined by determining signaling entity 33. For example, if the signaling entity is fluorescent, then fluorescence microscopy can be used to determine the signaling entity. In some embodiments, imaging of a sample to determine the signaling entity may be used at relatively

high resolutions, and in some cases, super-resolution imaging techniques (e.g., resolutions better than the wavelength of visible light or the diffraction limit of light) may be used. Examples of super-resolution imaging techniques include STORM, or other techniques as discussed herein. In some cases, e.g., with certain super-resolution imaging techniques such as STORM, more than one image of the sample may be acquired.

More than one type of secondary nucleic acid probe may be applied to a cell or other sample. For example, a first secondary nucleic acid probe may be applied that can recognize a first read sequence, then it or its attached signaling entity may be inactivated or removed, and a second secondary nucleic acid probe may be applied that can recognize a second read sequence. This process may be repeated multiple times, each with a different secondary nucleic acid probe, e.g., to determine the read sequences that were present in the various primary nucleic acid probes. Thus, primary nucleic acids within the sample can be determined on the basis of the binding pattern of secondary nucleic acid probes.

For example, a first location within the cell or other sample may exhibit binding of a first secondary probe and a third secondary probe, but not the binding of a second or a fourth secondary probe, while a second location may exhibit a different pattern of binding of various secondary probes. The primary nucleic acid probe that the secondary probes are able to bind to or hybridize with may be determined by considering the pattern of binding of various secondary probes. For instance, referring to Fig. 3A, if a first secondary probe is able to determine read sequence 27, a second secondary probe is able to determine read sequence 28, and a third secondary probe is able to determine read sequence 29, then primary nucleic acid 25 may be determined through the binding of the first and third secondary probes (but not the second secondary probe), while primary nucleic acid 26 may be determined through the binding of the first and second secondary probes (but not the third secondary probe). Similarly, if it is known that first probe 21 contains target sequence 25 while second probe 22 contains target sequence 26, then nucleic acids 11 and 12 may also be determined within the sample, e.g., spatially, based on the binding pattern of the various secondary nucleic acid probes. In addition, it should be noted that due to the presence of more than one read sequence on the primary nucleic acid probes, even though first probe 21 and second probe 22 contains a common read sequence (read sequence 27), these probes may be distinguished in the sample due to the different binding patterns of the various secondary nucleic acid probes.

In certain embodiments, this pattern of binding or hybridization of the secondary nucleic acid probes may be converted into a "codeword." In this example, for instance, the codewords are "101" and "110" for first probe 21 and second probe 22, respectively, where a

value of 1 represents binding and a value of 0 represents no binding. The codewords may also have longer lengths in other embodiments; only three probes are shown here for clarity purposes only. A codeword can be directly related to a specific target nucleic acid sequence of the primary nucleic acid probe. Accordingly, different primary nucleic acid probes may match certain codewords, which can then be used to identify the different targets of the primary nucleic acid probes based on the binding patterns of the secondary probes, even if in some cases, there is overlap in the read sequences of different secondary probes, e.g., as was shown in Fig. 3A. However, if no binding is evident (e.g., for nucleic acid 13), then the codeword would be "000" in this example.

The values in each codeword can also be assigned in different fashions in some embodiments. For example, a value of 0 could represent binding while a value of 1 represents no binding. Similarly, a value of 1 could represent binding of a secondary nucleic acid probe with one type of signaling entity while a value of 0 could represent binding of a secondary nucleic acid probe with another type of distinguishable signaling entity. These signaling entities could be distinguished, for example, via different colors of fluorescence. In some cases, values in codewords need not be confined to 0 and 1. The values could also be drawn from larger alphabets, such as ternary (e.g., 0, 1, and 2) or quaternary (e.g., 0, 1, 2, and 3) systems. Each different value could, for example, be represented by a different distinguishable signaling entity, including (in some cases) one value that may be represented by the absence of signal.

The codewords for each target may be assigned sequentially, or may be assigned at random. For instance, referring to Fig. 3A, a first nucleic acid target may be assigned to 101, while a second nucleic acid target may be assigned to 110. In addition, in some embodiments, the codewords may be assigned using an error-detection system or an error-correcting system, such as a Hamming system, a Golay code, or an extended Hamming system (or a SECDED system, i.e., single error correction, double error detection). Generally speaking, such systems can be used to identify where errors have occurred, and in some cases, such systems can also be used to correct the errors and determine what the correct codeword should have been. For example, a codeword such as 001 may be detected as invalid and corrected using such a system to 101, e.g., if 001 is not previously assigned to a different target sequence. A variety of different error-correcting codes can be used, many of which have previously been developed for use within the computer industry; however, such error-correcting systems have not typically been used within biological systems. Additional examples of such error-correcting codes are discussed in more detail below.

It should also be understood that all possible codewords in a code need not be used in some cases. For example, in some embodiments, codewords that are not used can serve as negative controls. Similarly, in some embodiments, some codewords can be left out because they are more prone to errors in measurement than other codewords. For example, in some implementations, reading a codeword with more values of '1' might be more error-prone that reading a codeword with fewer values of '1.'

It should be understood that the above description is an example of one embodiment of the invention, and that primary and secondary nucleic acid probes are not necessary in all embodiments. For example, in some embodiments, a series of nucleic acid probes containing signaling entities are used to determine nucleic acids within a cell or other sample, without necessarily requiring secondary probes.

For example, turning now to Fig. 3B, nucleic acids 11, 12, and 13 are exposed to different rounds of probes 21, 22, 23, 24, etc. in this example. These probes may each contain a target sequence that can recognize a nucleic acid (e.g., a sequence within nucleic acid 11 or 12). These probes may each target the same nucleic acid, but different regions of the nucleic acid. In addition, some or all of the probes may contain one or more signaling entities, e.g., signaling entity 29 on probe 21. For example, the signaling entity may be a fluorescent entity attached to the probe, or a certain sequence of nucleic acids that can be determined in some fashion.

The first round of probes (e.g. probe 21 and probe 22) may be applied to the cell or other sample. Probe 21 may be allowed to bind to nucleic acid 11 via target sequence 25. Such binding can be determined by determining signaling entity 29. For example, if the signaling entity is fluorescent, then fluorescence microscopy can be used to determine the signaling entity, e.g., spatially within the cell or other sample. In some but not all embodiments, imaging of a sample to determine the signaling entity may be used at relatively high resolutions, and in some cases, super-resolution imaging techniques may be used. Other, different probes may be present as well; for instance, probe 22 containing target sequence 26 may bind to nucleic acid 12, and be determined via signaling entity 29 within probe 22. These may occur, e.g., sequentially or simultaneously. Optionally, probes 21 and 22 may also be removed or inactivated, e.g., between application of different rounds of probes.

Next, a second round of probes (e.g., probe 23) is applied to the sample. In this example, probe 23 is able to bind to nucleic acid 11 via a targeting region, although there is no probe in the second round that is able to bind to nucleic acid 12. Binding of the probes is

allowed to occur as discussed above, and determination of binding may occur via signaling entities. These signaling entities may be the same or different as from the first round of probes. This process may be repeated any number of times with different probes. For example, as is shown in Fig. 3B, round 2 contains probes able to bind to nucleic acid 11, while round 3 contains probes able to bind to nucleic acid 12.

In certain embodiments, each round of binding or hybridization of nucleic acid probes may be converted into a "codeword." In this example, using probes 21, 22, 23, and 24, the codewords 101 or 110 could be formed, where 1 represents binding and 0 represents no binding and the first position corresponds to the binding of probes 21 or 22 while the second position corresponds to the binding of probes 22, and the third position corresponds to the binding of probe 24. A codeword of 000 would represent no binding, e.g., as shown with nucleic acid 13 in this example. A codeword can be directly related to a specific target nucleic acid sequence of the nucleic acid probes, by designing appropriate nucleic acid probes. Thus, for example, 110 may correspond to a first target nucleic acid 12 (e.g., the first and second round of nucleic acid probes containing probes able to target nucleic acid 11, and these probes may target the same or different regions of nucleic acid 11) while 101 may correspond to a second target nucleic acid (e.g., the first and third round of nucleic acid probes containing probes able to target nucleic acid 12, and these probes may target the same or different regions of nucleic acid 12). In addition, it should be noted that each round of probes may contain the same, or different signaling entities as other probes in the same round, and/or other probes in different rounds. For instance, in one set of embodiments, only one signaling entity is used in all of the rounds of probes.

Similar to the above, the codewords for each target may be assigned sequentially, or may be assigned at random. The codewords may be assigned within a code space in some embodiments using an error-detection or an error-correcting system, such as a Hamming system, a Golay code, or an extended Hamming system or a SECDED system (single error correction, double error detection). Generally speaking, such error-correction systems can be used to identify where errors have occurred, and in some cases, such systems can also be used to correct the errors and determine what the correct codeword should have been.

Similar to the above, the values at each position in the codeword can be arbitrarily assigned in certain embodiments to binding or non-binding of probes that contain more than one distinguishable signaling entity.

In some cases, the nucleic acid probes may be formed into "pools" or groups of nucleic acids that share a common feature. For example, probes to all targets with codewords

that contain a 1 in the first position, e.g. 110 and 101 but not 011, may comprise one pool while probes to all targets that contain a 1 in the second position, e.g. 110 and 011 but not 101, may comprise another pool. See also Fig. 1C. In some cases, a nucleic acid probe may be a member of more than one group or pool. Members of a nucleic acid pool may also contain features in addition to target sequences, read sequences, and or signaling entities that allow them to be distinguished from other groups. These features may be short nucleic acid sequences that are used for the amplification, production, or separation of these sequences. The nucleic acid probes of each group may be applied to a sample, e.g., sequentially, as discussed herein.

Thus, in some aspects, the present invention is generally directed to systems and methods for determining nucleic acids within a cell or other sample. The sample may include a cell culture, a suspension of cells, a biological tissue, a biopsy, an organism, or the like. The sample may also be cell-free but nevertheless contain nucleic acids. If the sample contains a cell, the cell may be a human cell, or any other suitable cell, e.g., a mammalian cell, a fish cell, an insect cell, a plant cell, or the like. More than one cell may be present in some cases.

The nucleic acids to be determined may be, for example, DNA, RNA, or other nucleic acids that are present within a cell (or other sample). The nucleic acids may be endogenous to the cell, or added to the cell. For instance, the nucleic acid may be viral, or artificially created. In some cases, the nucleic acid to be determined may be expressed by the cell. The nucleic acid is RNA in some embodiments. The RNA may be coding and/or non-coding RNA. Non-limiting examples of RNA that may be studied within the cell include mRNA, siRNA, rRNA, miRNA, tRNA, lncRNA, snoRNAs, snRNAs, exRNAs, piRNAs, or the like.

In some cases, a significant portion of the nucleic acid within the cell may be studied. For instance, in some cases, enough of the RNA present within a cell may be determined so as to produce a partial or complete transcriptome of the cell. In some cases, at least 4 types of mRNAs are determined within a cell, and in some cases, at least 3, at least 4, at least 7, at least 8, at least 12, at least 14, at least 15, at least 16, at least 22, at least 30, at least 31, at least 32, at least 50, at least 63, at least 64, at least 72, at least 75, at least 100, at least 127, at least 128, at least 140, at least 255, at least 256, at least 500, at least 1,000, at least 1,500, at least 2,000, at least 2,500, at least 3,000, at least 4,000, at least 5,000, at least 7,500, at least 10,000, at least 12,000, at least 15,000, at least 20,000, at least 25,000, at least 30,000, at least 40,000, at least 50,000, at least 75,000, or at least 100,000 types of mRNAs may be determined within a cell.

In some cases, the transcriptome of a cell may be determined. It should be understood that the transriptome generally encompasses all RNA molecules produced within a cell, not just mRNA. Thus, for instance, the transcriptome may also include rRNA, tRNA, siRNA, etc. In some embodiments, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or 100% of the transcriptome of a cell may be determined.

The determination of one or more nucleic acids within the cell or other sample may be qualitative and/or quantitative. In addition, the determination may also be spatial, e.g., the position of the nucleic acid within the cell or other sample may be determined in two or three dimensions. In some embodiments, the positions, number, and/or concentrations of nucleic acids within the cell (or other sample) may be determined.

In some cases, a significant portion of the genome of a cell may be determined. The determined genomic segments may be continuous or interspersed on the genome. For example, in some cases, at least 4 genomic segments are determined within a cell, and in some cases, at least 3, at least 4, at least 7, at least 8, at least 12, at least 14, at least 15, at least 16, at least 22, at least 30, at least 31, at least 32, at least 50, at least 63, at least 64, at least 72, at least 75, at least 100, at least 127, at least 128, at least 140, at least 255, at least 256, at least 500, at least 1,000, at least 1,500, at least 2,000, at least 2,500, at least 3,000, at least 4,000, at least 5,000, at least 7,500, at least 10,000, at least 12,000, at least 15,000, at least 20,000, at least 25,000, at least 30,000, at least 40,000, at least 50,000, at least 75,000, or at least 100,000 genomic segments may be determined within a cell.

In some cases, the entire genome of a cell may be determined. It should be understood that the genome generally encompasses all DNA molecules produced within a cell, not just chromosome DNA. Thus, for instance, the genome may also include, in some cases, mitochondria DNA, chloroplast DNA, plasmid DNA, etc. In some embodiments, at least about 5%, at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or 100% of the genome of a cell may be determined.

As discussed herein, a variety of nucleic acid probes may be used to determine one or more nucleic acids within a cell or other sample. The probes may comprise nucleic acids (or entities that can hybridize to a nucleic acid, e.g., specifically) such as DNA, RNA, LNA (locked nucleic acids), PNA (peptide nucleic acids), or combinations thereof. In some cases,

additional components may also be present within the nucleic acid probes, e.g., as discussed below. Any suitable method may be used to introduce nucleic acid probes into a cell.

For example, in some embodiments, the cell is fixed prior to introducing the nucleic acid probes, e.g., to preserve the positions of the nucleic acids within the cell. Techniques for fixing cells are known to those of ordinary skill in the art. As non-limiting examples, a cell may be fixed using chemicals such as formaldehyde, paraformaldehyde, glutaraldehyde, ethanol, methanol, acetone, acetic acid, or the like. In one embodiment, a cell may be fixed using Hepes-glutamic acid buffer-mediated organic solvent (HOPE).

The nucleic acid probes may be introduced into the cell (or other sample) using any suitable method. In some cases, the cell may be sufficiently permeabilized such that the nucleic acid probes may be introduced into the cell by flowing a fluid containing the nucleic acid probes around the cells. In some cases, the cells may be sufficiently permeabilized as part of a fixation process; in other embodiments, cells may be permeabilized by exposure to certain chemicals such as ethanol, methanol, Triton, or the like. In addition, in some embodiments, techniques such as electroporation or microinjection may be used to introduce nucleic acid probes into a cell or other sample.

Certain aspects of the present invention are generally directed to nucleic acid probes that are introduced into a cell (or other sample). The probes may comprise any of a variety of entities that can hybridize to a nucleic acid, typically by Watson-Crick base pairing, such as DNA, RNA, LNA, PNA, etc., depending on the application. The nucleic acid probe typically contains a target sequence that is able to bind to at least a portion of a target nucleic acid, in some cases specifically. When introduced into a cell or other system, the target system may be able to bind to a specific target nucleic acid (e.g., an mRNA, or other nucleic acids as discussed herein). In some cases, the nucleic acid probes may be determined using signaling entities (e.g., as discussed below), and/or by using secondary nucleic acid probes able to bind to the nucleic acid probes (i.e., to primary nucleic acid probes). The determination of such nucleic acid probes is discussed in detail below.

In some cases, more than one type of (primary) nucleic acid probe may be applied to a sample, e.g., simultaneously. For example, there may be at least 2, at least 5, at least 10, at least 25, at least 50, at least 75, at least 100, at least 300, at least 1,000, at least 3,000, at least 10,000, or at least 30,000 distinguishable nucleic acid probes that are applied to a sample, e.g., simultaneously or sequentially.

The target sequence may be positioned anywhere within the nucleic acid probe (or primary nucleic acid probe or encoding nucleic acid probe). The target sequence may contain

a region that is substantially complementary to a portion of a target nucleic acid. In some cases, the portions may be at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% complementary. In some cases, the target sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 50, at least 60, at least 65, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides in length. In some cases, the target sequence may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the target sequence may have a length of between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc. Typically, complementarity is determined on the basis of Watson-Crick nucleotide base pairing.

The target sequence of a (primary) nucleic acid probe may be determined with reference to a target nucleic acid suspected of being present within a cell or other sample. For example, a target nucleic acid to a protein may be determined using the protein's sequence, by determining the nucleic acids that are expressed to form the protein. In some cases, only a portion of the nucleic acids encoding the protein are used, e.g., having the lengths as discussed above. In addition, in some cases, more than one target sequence that can be used to identify a particular target may be used. For instance, multiple probes can be used, sequentially and/or simultaneously, that can bind to or hybridize to different regions of the same target. Hybridization typically refers to an annealing process by which complementary single-stranded nucleic acids associate through Watson-Crick nucleotide base pairing (e.g., hydrogen bonding, guanine-cytosine and adenine-thymine) to form double-stranded nucleic acid.

In some embodiments, a nucleic acid probe, such as a primary nucleic acid probe, may also comprise one or more "read" sequences. However, it should be understood that read sequences are not necessary in all cases. In some embodiments, the nucleic acid probe may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 or more, 20 or more, 32 or more, 40 or more, 50 or more, 64 or more, 75 or more, 100 or more, 128 or more read

sequences. The read sequences may be positioned anywhere within the nucleic acid probe. If more than one read sequence is present, the read sequences may be positioned next to each other, and/or interspersed with other sequences.

The read sequences, if present, may be of any length. If more than one read sequence is used, the read sequences may independently have the same or different lengths. For instance, the read sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 50, at least 60, at least 65, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides in length. In some cases, the read sequence may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the read sequence may have a length of between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc.

The read sequence may be arbitrary or random in some embodiments. In certain cases, the read sequences are chosen so as to reduce or minimize homology with other components of the cell or other sample, e.g., such that the read sequences do not themselves bind to or hybridize with other nucleic acids suspected of being within the cell or other sample. In some cases, the homology may be less than 10%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, or less than 1%. In some cases, there may be a homology of less than 20 basepairs, less than 18 basepairs, less than 15 basepairs, less than 14 basepairs, less than 13 basepairs, less than 12 basepairs, less than 11 basepairs, or less than 10 basepairs. In some cases, the basepairs are sequential.

In one set of embodiments, a population of nucleic acid probes may contain a certain number of read sequences, which may be less than the number of targets of the nucleic acid probes in some cases. Those of ordinary skill in the art will be aware that if there is one signaling entity and n read sequences, then in general 2^n-1 different nucleic acid targets may be uniquely identified. However, not all possible combinations need be used. For instance, a population of nucleic acid probes may target 12 different nucleic acid sequences, yet contain no more than 8 read sequences. As another example, a population of nucleic acids may target 140 different nucleic acid species, yet contain no more than 16 read sequences. Different

nucleic acid sequence targets may be separately identified by using different combinations of read sequences within each probe. For instance, each probe may contain 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, etc. or more read sequences. In some cases, a population of nucleic acid probes may each contain the same number of read sequences, although in other cases, there may be different numbers of read sequences present on the various probes.

As a non-limiting example, a first nucleic acid probe may contain a first target sequence, a first read sequence, and a second read sequence, while a second, different nucleic acid probe may contain a second target sequence, the same first read sequence, but a third read sequence instead of the second read sequence. Such probes may thereby be distinguished by determining the various read sequences present or associated with a given probe or location, as discussed herein.

In addition, the nucleic acid probes (and their corresponding, complimentary sites on the encoding probes), in certain embodiments, may be made using only 2 or only 3 of the 4 bases, such as leaving out all the "G"s or leaving out all of the "C"s within the probe. Sequences lacking either "G"s or "C"s may form very little secondary structure in certain embodiments, and can contribute to more uniform, faster hybridization.

In some embodiments, the nucleic acid probe may contain a signaling entity. It should be understood that signaling entities are not required in all cases, however; for instance, the nucleic acid probe may be determined using secondary nucleic acid probes in some embodiments, as is discussed in additional detail below. Examples of signaling entities that can be used are also discussed in more detail below.

Other components may also be present within a nucleic acid probe as well. For example, in one set of embodiments, one or more primer sequences may be present, e.g., to allow for enzymatic amplification of probes. Those of ordinary skill in the art will be aware of primer sequences suitable for applications such as amplification (e.g., using PCR or other suitable techniques). Many such primer sequences are available commercially. Other examples of sequences that may be present within a primary nucleic acid probe include, but are not limited to promoter sequences, operons, identification sequences, nonsense sequences, or the like.

Typically, a primer is a single-stranded or partially double-stranded nucleic acid (e.g., DNA) that serves as a starting point for nucleic acid synthesis, allowing polymerase enzymes such as nucleic acid polymerase to extend the primer and replicate the complementary strand. A primer is (e.g., is designed to be) complementary to and to hybridize to a target nucleic acid. In some embodiments, a primer is a synthetic primer. In some embodiments, a primer

is a non-naturally-occurring primer. A primer typically has a length of 10 to 50 nucleotides. For example, a primer may have a length of 10 to 40, 10 to 30, 10 to 20, 25 to 50, 15 to 40, 15 to 30, 20 to 50, 20 to 40, or 20 to 30 nucleotides. In some embodiments, a primer has a length of 18 to 24 nucleotides.

In addition, the components of the nucleic acid probe may be arranged in any suitable order. For instance, in one embodiment, the components may be arranged in a nucleic acid probe as: primer—read sequences—targeting sequence—read sequences—reverse primer. The "read sequences" in this structure may each contain any number (including 0) of read sequences, so long as at least one read sequence is present in the probe. Non-limiting example structures include primer—targeting sequence—read sequences—reverse primer, primer—read sequences—targeting sequence—reverse primer, targeting sequence—primer—targeting sequence—read sequences—reverse primer, targeting sequence—primer—read sequences—targeting sequence—reverse primer, primer—target sequence—read sequences—targeting sequence—reverse primer, targeting sequence—primer—read sequence—reverse primer, targeting sequence—read sequence—primer, read sequence—targeting sequence—primer, read sequence—primer—targeting sequence—reverse primer, etc. In addition, the reverse primer is optional in some embodiments, including in all of the above-described examples.

After introduction of the nucleic acid probes into a cell or other sample, the nucleic acid probes may be directly determined by determining signaling entities (if present), and/or the nucleic acid probes may be determined by using one or more secondary nucleic acid probes, in accordance with certain aspects of the invention. As mentioned, in some cases, the determination may be spatial, e.g., in two or three dimensions. In addition, in some cases, the determination may be quantitative, e.g., the amount or concentration of a primary nucleic acid probe (and of a target nucleic acid) may be determined. Additionally, the secondary probes may comprise any of a variety of entities able to hybridize a nucleic acid, e.g., DNA, RNA, LNA, and/or PNA, etc., depending on the application. Signaling entities are discussed in more detail below.

A secondary nucleic acid probe may contain a recognition sequence able to bind to or hybridize with a read sequence of a primary nucleic acid probe. In some cases, the binding is specific, or the binding may be such that a recognition sequence preferentially binds to or hybridizes with only one of the read sequences that are present. The secondary nucleic acid probe may also contain one or more signaling entities. If more than one secondary nucleic acid probe is used, the signaling entities may be the same or different.

The recognition sequences may be of any length, and multiple recognition sequences may be of the same or different lengths. If more than one recognition sequence is used, the recognition sequences may independently have the same or different lengths. For instance, the recognition sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, or at least 50 nucleotides in length. In some cases, the recognition sequence may be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the recognition sequence may have a length of between 10 and 30, between 20 and 40, or between 25 and 35 nucleotides, etc. In one embodiment, the recognition sequence is of the same length as the read sequence. In addition, in some cases, the recognition sequence may be at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 100% complementary to a read sequence of the primary nucleic acid probe.

As mentioned, in some cases, the secondary nucleic acid probe may comprise one or more signaling entities. Examples of signaling entities are discussed in more detail below.

As discussed, in certain aspects of the invention, nucleic acid probes are used that contain various "read sequences." For example, a population of primary nucleic acid probes may contain certain "read sequences" which can bind certain of the secondary nucleic acid probes, and the locations of the primary nucleic acid probes are determined within the sample using secondary nucleic acid probes, e.g., which comprise a signaling entity. As mentioned, in some cases, a population of read sequences may be combined in various combinations to produce different nucleic acid probes, e.g., such that a relatively small number of read sequences may be used to produce a relatively large number of different nucleic acid probes.

Thus, in some cases, a population of primary nucleic acid probes (or other nucleic acid probes) may each contain a certain number of read sequences, some of which are shared between different primary nucleic acid probes such that the total population of primary nucleic acid probes may contain a certain number of read sequences. A population of nucleic acid probes may have any suitable number of read sequences. For example, a population of primary nucleic acid probes may have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 etc. read sequences. More than 20 are also possible in some embodiments. In addition, in some cases, a population of nucleic acid probes may, in total, have 1 or more, 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or

more, 11 or more, 12 or more, 13 or more, 14 or more, 15 or more, 16 or more, 20 or more, 24 or more, 32 or more, 40 or more, 50 or more, 60 or more, 64 or more, 100 or more, 128 or more, etc. of possible read sequences present, although some or all of the probes may each contain more than one read sequence, as discussed herein. In addition, in some embodiments, the population of nucleic acid probes may have no more than 100, no more than 80, no more than 64, no more than 60, no more than 50, no more than 40, no more than 32, no more than 24, no more than 20, no more than 16, no more than 15, no more than 14, no more than 13, no more than 12, no more than 11, no more than 10, no more than 9, no more than 8, no more than 7, no more than 6, no more than 5, no more than 4, no more than 3, or no more than two read sequences present. Combinations of any of these are also possible, e.g., a population of nucleic acid probes may comprise between 10 and 15 read sequences in total.

As a non-limiting example of an approach to combinatorially producing a relatively large number of nucleic acid probes from a relatively small number of read sequences, in a population of 6 different types of nucleic acid probes, each comprising one or more read sequences, the total number of read sequences within the population may be no greater than 4. It should be understood that although 4 read sequences are used in this example for ease of explanation, in other embodiments, larger numbers of nucleic acid probes may be realized, for example, using 5, 8, 10, 16, 32, etc. or more read sequences, or any other suitable number of read sequences described herein, depending on the application. Referring now to Fig. 4A, if each of the primary nucleic acid probes contains two different read sequences, then by using 4 such read sequences (A, B, C, and D), up to 6 probes may be separately identified. It should be noted that in this example, the ordering of read sequences on a nucleic acid probe is not essential, i.e., "AB" and "BA" may be treated as being synonymous (although in other embodiments, the ordering of read sequences may be essential and "AB" and "BA" may not necessarily be synonymous). Similarly, if 5 read sequences are used (A, B, C, D, and E) in the population of primary nucleic acid probes, up to 10 probes may be separately identified, as is shown in Fig. 4B. For example, one of ordinary skill in the art would understand that, for k read sequences in a population with n read sequences on each probe, up to $\binom{n}{k}$ different probes may be produced, assuming that the ordering of read sequences is not essential; because not all of the probes need to have the same number of read sequences and not all combinations of read sequences need to be used in every embodiment, either more or less than this number of different probes may also be used in certain embodiments. In addition, it should also be understood that the number of read sequences on each probe need not be

identical in some embodiments. For instance example, some probes may contain 2 read sequences while other probes may contain 3 read sequences.

In some aspects, the read sequences and/or the pattern of binding of nucleic acid probes within a sample may be used to define an error-detecting and/or an error-correcting code, for example, to reduce or prevent misidentification or errors of the nucleic acids, e.g., as was discussed with reference to Fig. 3. Thus, for example, if binding is indicated (e.g., as determined using a signaling entity), then the location may be identified with a "1"; conversely, if no binding is indicated, then the location may be identified with a "0" (or vice versa, in some cases). Multiple rounds of binding determinations, e.g., using different nucleic acid probes, can then be used to create a "codeword," e.g., for that spatial location. In some embodiments, the codeword may be subjected to error detection and/or correction. For instance, the codewords may be organized such that, if no match is found for a given set of read sequences or binding pattern of nucleic acid probes, then the match may be identified as an error, and optionally, error correction may be applied sequences to determine the correct target for the nucleic acid probes. In some cases, the codewords may have fewer "letters" or positions that the total number of nucleic acids encoded by the codewords, e.g. where each codeword encodes a different nucleic acid.

Such error-detecting and/or the error-correction code may take a variety of forms. A variety of such codes have previously been developed in other contexts such as the telecommunications industry, such as Golay codes or Hamming codes. In one set of embodiments, the read sequences or binding patterns of the nucleic acid probes are assigned such that not every possible combination is assigned.

For example, if 4 read sequences are possible and a primary nucleic acid probe contains 2 read sequences, then up to 6 primary nucleic acid probes could be identified; but the number of primary nucleic acid probes used may be less than 6. Similarly, for k read sequences in a population with n read sequences on each primary nucleic acid probe, $\binom{n}{k}$ different probes may be produced, but the number of primary nucleic acid probes that are used may be any number more or less than $\binom{n}{k}$. In addition, these may be randomly assigned, or assigned in specific ways to increase the ability to detect and/or correct errors.

As another example, if multiple rounds of nucleic acid probes are used, the number of rounds may be arbitrarily chosen. If in each round, each target can give two possible outcomes, such as being detected or not being detected, up to 2^n different targets may be possible for n rounds of probes, but the number of nucleic acid targets that are actually used

may be any number less than 2^n. For example, if in each round, each target can give more than two possible outcomes, such as being detected in different color channels, more than 2^n (e.g. 3^n, 4^n ...) different targets may be possible for n rounds of probes. In some cases, the number of nucleic acid targets that are actually used may be any number less than this number. In addition, these may be randomly assigned, or assigned in specific ways to increase the ability to detect and/or correct errors.

For example, in one set of embodiments, the codewords or nucleic acid probes may be assigned within a code space such that the assignments are separated by a Hamming distance, which measures the number of incorrect "reads" in a given pattern that cause the nucleic acid probe to be misinterpreted as a different valid nucleic acid probe. In certain cases, the Hamming distance may be at least 2, at least 3, at least 4, at least 5, at least 6, or the like. In addition, in one set of embodiments, the assignments may be formed as a Hamming code, for instance, a Hamming(7, 4) code, a Hamming(15, 11) code, a Hamming(31, 26) code, a Hamming(63, 57) code, a Hamming(127, 120) code, etc. In another set of embodiments, the assignments may form a SECDED code, e.g., a SECDED(8,4) code, a SECDED(16,4) code, a SCEDED(16, 11) code, a SCEDED(22, 16) code, a SCEDED(39, 32) code, a SCEDED(72, 64) code, etc. In yet another set of embodiments, the assignments may form an extended binary Golay code, a perfect binary Golay code, or a ternary Golay code. In another set of embodiments, the assignments may represent a subset of the possible values taken from any of the codes described above.

For example, a code with the same error correcting properties of the SECDED code may be formed by using only binary words that contain a fixed number of '1' bits, such as 4, to encode the targets. In another set of embodiments, the assignments may represent a subset of the possible values taken from codes described above for the purpose of addressing asymmetric readout errors. For example, in some cases, a code in which the number of '1' bits may be fixed for all used binary words may eliminate the biased measurement of words with different numbers of '1's when the rate at which '0' bits are measured as '1's or '1' bits are measured as '0's are different.

Accordingly, in some embodiments, once the codeword is determined (e.g., as discussed herein), the codeword may be compared to the known nucleic acid codewords. If a match is found, then the nucleic acid target can be identified or determined. If no match is found, then an error in the reading of the codeword may be identified. In some cases, error correction can also be applied to determine the correct codeword, and thus resulting in the correct identity of the nucleic acid target. In some cases, the codewords may be selected such

that, assuming that there is only one error present, only one possible correct codeword is available, and thus, only one correct identity of the nucleic acid target is possible. In some cases, this may also be generalized to larger codeword spacings or Hamming distances; for instance, the codewords may be selected such that if two, three, or four errors are present (or more in some cases), only one possible correct codeword is available, and thus, only one correct identity of the nucleic acid targets is possible.

The error-correcting code may be a binary error-correcting code, or it may be based on other numbering systems, e.g., ternary or quaternary error-correcting codes. For instance, in one set of embodiments, more than one type of signaling entity may be used and assigned to different numbers within the error-correcting code. Thus, as a non-limiting example, a first signaling entity (or more than one signaling entity, in some cases) may be assigned as "1" and a second signaling entity (or more than one signaling entity, in some cases) may be assigned as "2" (with "0" indicating no signaling entity present), and the codewords distributed to define a ternary error-correcting code. Similarly, a third signaling entity may additionally be assigned as "3" to make a quaternary error-correcting code, etc.

As discussed above, in certain aspects, signaling entities are determined, e.g., to determine nucleic acid probes and/or to create codewords. In some cases, signaling entities within a sample may be determined, e.g., spatially, using a variety of techniques. In some embodiments, the signaling entities may be fluorescent, and techniques for determining fluorescence within a sample, such as fluorescence microscopy or confocal microscopy, may be used to spatially identify the positions of signaling entities within a cell. In some cases, the positions of entities within the sample may be determined in two or even three dimensions. In addition, in some embodiments, more than one signaling entity may be determined at a time (e.g., signaling entities with different colors or emissions), and/or sequentially.

In addition, in some embodiments, a confidence level for the identified nucleic acid target may be determined. For example, the confidence level may be determined using a ratio of the number of exact matches to the number of matches having one or more one-bit errors. In some cases, only matches having a confidence ratio greater than a certain value may be used. For instance, in certain embodiments, matches may be accepted only if the confidence ratio for the match is greater than about 0.01, greater than about 0.03, greater than about 0.05, greater than about 0.1, greater than about 0.3, greater than about 0.5, greater than about 1, greater than about 3, greater than about 5, greater than about 10, greater than about 30, greater than about 50, greater than about 100, greater than about 300, greater than about

500, greater than about 1000, or any other suitable value. In addition, in some embodiments, matches may be accepted only if the confidence ratio for the identified nucleic acid target is greater than an internal standard or false positive control by about 0.01, about 0.03, about 0.05, about 0.1, about 0.3, about 0.5, about 1, about 3, about 5, about 10, about 30, about 50, about 100, about 300, about 500, about 1000, or any other suitable value

In some embodiments, the spatial positions of the entities (and thus, nucleic acid probes that the entities may be associated with) may be determined at relatively high resolutions. For instance, the positions may be determined at spatial resolutions of better than about 100 micrometers, better than about 30 micrometers, better than about 10 micrometers, better than about 3 micrometers, better than about 1 micrometer, better than about 800 nm, better than about 600 nm, better than about 500 nm, better than about 400 nm, better than about 300 nm, better than about 200 nm, better than about 100 nm, better than about 90 nm, better than about 80 nm, better than about 70 nm, better than about 60 nm, better than about 50 nm, better than about 40 nm, better than about 30 nm, better than about 20 nm, or better than about 10 nm, etc.

There are a variety of techniques able to determine or image the spatial positions of entities optically, e.g., using fluorescence microscopy. In some cases, the spatial positions may be determined at super resolutions, or at resolutions better than the wavelength of light or the diffraction limit. Non-limiting examples include STORM (stochastic optical reconstruction microscopy), STED (stimulated emission depletion microscopy), NSOM (Near-field Scanning Optical Microscopy), 4Pi microscopy, SIM (Structured Illumination Microscopy), SMI (Spatially Modulated Illumination) microscopy, RESOLFT (Reversible Saturable Optically Linear Fluorescence Transition Microscopy), GSD (Ground State Depletion Microscopy), SSIM (Saturated Structured-Illumination Microscopy), SPDM (Spectral Precision Distance Microscopy), Photo-Activated Localization Microscopy (PALM), Fluorescence Photoactivation Localization Microscopy (FPALM), LIMON (3D Light Microscopical Nanosizing Microscopy), Super-resolution optical fluctuation imaging (SOFI), or the like. See, e.g., U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*; or Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*, each incorporated herein by reference in their entireties.

As an illustrative non-limiting example, in one set of embodiments, the sample may be imaged with a high numerical aperture, oil immersion objective with 100X magnification and light collected on an electron-multiplying CCD camera. In another example, the sample could be imaged with a high numerical aperture, oil immersion lens with 40X magnification and light collected with a wide-field scientific CMOS camera. With different combinations of objectives and cameras, a single field of view may correspond to no less than 40 x 40 microns, 80 x 80 microns, 120 x 120 microns, 240 x 240 microns, 340 x 340 microns, or 500 x 500 microns, etc. in various non-limiting embodiments. Similarly, a single camera pixel may correspond, in some embodiments, to regions of the sample of no less than 80x80 nm, 120x120 nm, 160x160 nm, 240x240 nm, or 300x300 nm, etc. In another example, the sample may be imaged with a low numerical aperture, air lens with 10X magnification and light collected with a sCMOS camera. In additional embodiments, the sample may be optically sectioned by illuminating it via a single or multiple scanned diffraction limited foci generated either by scanning mirrors or a spinning disk and the collected passed through a single or multiple pinholes. In another embodiment, the sample may also be illuminated via thin sheet of light generated via any one of multiple methods known to those versed in the art.

In one embodiment, the sample may be illuminated by single Gaussian mode laser lines. In some embodiments, the illumination profiled may be flattened by passing these laser lines through a multimode fiber that is vibrated via piezo-electric or other mechanical means. In some embodiments, the illumination profile may be flattened by passing single-mode, Gaussian beams through a variety of refractive beam shapers, such as the piShaper or a series of stacked Powell lenses. In yet another set of embodiments, the Gaussian beams may be passed through a variety of different diffusing elements, such as ground glass or engineered diffusers, which may be spun in some cases at high speeds to remove residual laser speckle. In yet another embodiment, laser illumination may be passed through a series of lenslet arrays to produce overlapping images of the illumination that approximate a flat illumination field.

In some embodiments, the centroids of the spatial positions of the entities may be determined. For example, a centroid of a signaling entity may be determined within an image or series of images using image analysis algorithms known to those of ordinary skill in the art. In some cases, the algorithms may be selected to determine non-overlapping single emitters and/or partially overlapping single emitters in a sample. Non-limiting examples of suitable techniques include a maximum likelihood algorithm, a least squares algorithm, a

Bayesian algorithm, a compressed sensing algorithm, or the like. Combinations of these techniques may also be used in some cases.

In addition, the signaling entity may be inactivated in some cases. For example, in some embodiments, a first secondary nucleic acid probe containing a signaling entity may be applied to a sample that can recognize a first read sequence, then the first secondary nucleic acid probe can be inactivated before a second secondary nucleic acid probe is applied to the sample. If multiple signaling entities are used, the same or different techniques may be used to inactivate the signaling entities, and some or all of the multiple signaling entities may be inactivated, e.g., sequentially or simultaneously.

Inactivation may be caused by removal of the signaling entity (e.g., from the sample, or from the nucleic acid probe, etc.), and/or by chemically altering the signaling entity in some fashion, e.g., by photobleaching the signaling entity, bleaching or chemically altering the structure of the signaling entity, e.g., by reduction, etc.). For instance, in one set of embodiments, a fluorescent signaling entity may be inactivated by chemical or optical techniques such as oxidation, photobleaching, chemically bleaching, stringent washing or enzymatic digestion or reaction by exposure to an enzyme, dissociating the signaling entity from other components (e.g., a probe), chemical reaction of the signaling entity (e.g., to a reactant able to alter the structure of the signaling entity) or the like. For instance, bleaching may occur by exposure to oxygen, reducing agents, or the signaling entity could be chemically cleaved from the nucleic acid probe and washed away via fluid flow.

In some embodiments, various nucleic acid probes (including primary and/or secondary nucleic acid probes) may include one or more signaling entities. If more than one nucleic acid probe is used, the signaling entities may each by the same or different. In certain embodiments, a signaling entity is any entity able to emit light. For instance, in one embodiment, the signaling entity is fluorescent. In other embodiments, the signaling entity may be phosphorescent, radioactive, absorptive, etc. In some cases, the signaling entity is any entity that can be determined within a sample at relatively high resolutions, e.g., at resolutions better than the wavelength of visible light or the diffraction limit. The signaling entity may be, for example, a dye, a small molecule, a peptide or protein, or the like. The signaling entity may be a single molecule in some cases. If multiple secondary nucleic acid probes are used, the nucleic acid probes may comprise the same or different signaling entities.

Non-limiting examples of signaling entities include fluorescent entities (fluorophores) or phosphorescent entities, for example, cyanine dyes (e.g., Cy2, Cy3, Cy3B, Cy5, Cy5.5,

Cy7, etc.), Alexa Fluor dyes, Atto dyes, photoswtichable dyes, photoactivatable dyes, fluorescent dyes, metal nanoparticles, semiconductor nanoparticles or "quantum dots", fluorescent proteins such as GFP (Green Fluorescent Protein), or photoactivabale fluorescent proteins, such as PAGFP, PSCFP, PSCFP2, Dendra, Dendra2, EosFP, tdEos, mEos2, mEos3, PAmCherry, PAtagRFP, mMaple, mMaple2, and mMaple3. Other suitable signaling entities are known to those of ordinary skill in the art. See, e.g., U.S. Pat. No. 7,838,302 or U.S. Pat. Apl. Ser. No. 61/979,436, each incorporated herein by reference in its entirety.

In one set of embodiments, the signaling entity may be attached to an oligonucleotide sequence via a bond that can be cleaved to release the signaling entity. In one set of embodiments, a fluorophore may be conjugated to an oligonucleotide via a cleavable bond, such as a photocleavable bond. Non-limiting examples of photocleavable bonds include, but are not limited to, 1-(2-nitrophenyl)ethyl, 2-nitrobenzyl, biotin phosphoramidite, acrylic phosphoramidite, diethylaminocoumarin, 1-(4,5-dimethoxy-2-nitrophenyl)ethyl, cyclo-dodecyl (dimethoxy-2-nitrophenyl)ethyl, 4-aminomethyl-3-nitrobenzyl, (4-nitro-3-(1-chlorocarbonyloxyethyl)phenyl)methyl-S-acetylthioic acid ester, (4-nitro-3-(1-thlorocarbonyloxyethyl)phenyl)methyl-3-(2-pyridyldithiopropionic acid) ester, 3-(4,4'-dimethoxytrityl)-1-(2-nitrophenyl)-propane-1,3-diol-[2- cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-trifluoroacetylcaproamidomethyl)phenyl]-ethyl-[2-cyano-ethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-(4,4'-dimethoxytrityloxy)butyramidomethyl)phenyl]-ethyl-[2-cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-(N-(4,4'-dimethoxytrityl))-biotinamidocaproamido-methyl)phenyl]-ethyl-[2-cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, or similar linkers. In another set of embodiments, the fluorophore may be conjugated to an oligonucleotide via a disulfide bond. The disulfide bond may be cleaved by a variety of reducing agents such as, but not limited to, dithiothreitol, dithioerythritol, beta-mercaptoethanol, sodium borohydride, thioredoxin, glutaredoxin, trypsinogen, hydrazine, diisobutylaluminum hydride, oxalic acid, formic acid, ascorbic acid, phosphorous acid, tin chloride, glutathione, thioglycolate, 2,3-dimercaptopropanol, 2-mercaptoethylamine, 2-aminoethanol, tris(2-carboxyethyl)phosphine, bis(2-mercaptoethyl) sulfone, N,N'-dimethyl-N,N'-bis(mercaptoacetyl)hydrazine, 3-mercaptoproptionate, dimethylformamide, thiopropyl-agarose, tri-*n*-butylphosphine, cysteine, iron sulfate, sodium sulfite, phosphite, hypophosphite, phosphorothioate, or the like, and/or combinations of any of these. In another embodiment, the fluorophore may be conjugated to an oligonucleotide via one or more phosphorothioate modified nucleotides in which the sulfur modification replaces the bridging and/or non-bridging oxygen. The fluorophore may be

cleaved from the oligonucleotide, in certain embodiments, via addition of compounds such as but not limited to iodoethanol, iodine mixed in ethanol, silver nitrate, or mercury chloride. In yet another set of embodiments, the signaling entity may be chemically inactivated through reduction or oxidation. For example, in one embodiment, a chromophore such as Cy5 or Cy7 may be reduced using sodium borohydride to a stable, non-fluorescence state. In still another set of embodiments, a fluorophore may be conjugated to an oligonucleotide via an azo bond, and the azo bond may be cleaved with 2-[(2-N-arylamino)phenylazo]pyridine. In yet another set of embodiments, a fluorophore may be conjugated to an oligonucleotide via a suitable nucleic acid segment that can be cleaved upon suitable exposure to DNAse, e.g., an exodeoxyribonuclease or an endodeoxyribonuclease. Examples include, but are not limited to, deoxyribonuclease I or deoxyribonuclease II. In one set of embodiments, the cleavage may occur via a restriction endonuclease. Non-limiting examples of potentially suitable restriction endonucleases include BamHI, BsrI, NotI, XmaI, PspAI, DpnI, MboI, MnlI, Eco57I, Ksp632I, DraIII, AhaII, SmaI, MluI, HpaI, ApaI, BclI, BstEII, TaqI, EcoRI, SacI, HindII, HaeII, DraII, Tsp509I, Sau3AI, PacI, etc. Over 3000 restriction enzymes have been studied in detail, and more than 600 of these are available commercially. In yet another set of embodiments, a fluorophore may be conjugated to biotin, and the oligonucleotide conjugated to avidin or streptavidin. An interaction between biotin and avidin or streptavidin allows the fluorophore to be conjugated to the oligonucleotide, while sufficient exposure to an excess of addition, free biotin could "outcompete" the linkage and thereby cause cleavage to occur. In addition, in another set of embodiments, the probes may be removed using corresponding "toe-hold-probes," which comprise the same sequence as the probe, as well as an extra number of bases of homology to the encoding probes (e.g., 1-20 extra bases, for example, 5 extra bases). These probes may remove the labeled readout probe through a strand-displacement interaction.

As used herein, the term "light" generally refers to electromagnetic radiation, having any suitable wavelength (or equivalently, frequency). For instance, in some embodiments, the light may include wavelengths in the optical or visual range (for example, having a wavelength of between about 400 nm and about 700 nm, i.e., "visible light"), infrared wavelengths (for example, having a wavelength of between about 300 micrometers and 700 nm), ultraviolet wavelengths (for example, having a wavelength of between about 400 nm and about 10 nm), or the like. In certain cases, as discussed in detail below, more than one entity may be used, i.e., entities that are chemically different or distinct, for example,

structurally. However, in other cases, the entities may be chemically identical or at least substantially chemically identical.

In one set of embodiments, the signaling entity is "switchable," i.e., the entity can be switched between two or more states, at least one of which emits light having a desired wavelength. In the other state(s), the entity may emit no light, or emit light at a different wavelength. For instance, an entity may be "activated" to a first state able to produce light having a desired wavelength, and "deactivated" to a second state not able to emit light of the same wavelength. An entity is "photoactivatable" if it can be activated by incident light of a suitable wavelength. As a non-limiting example, Cy5, can be switched between a fluorescent and a dark state in a controlled and reversible manner by light of different wavelengths, i.e., 633 nm (or 642nm, 647nm, 656 nm) red light can switch or deactivate Cy5 to a stable dark state, while 405 nm green light can switch or activate the Cy5 back to the fluorescent state. In some cases, the entity can be reversibly switched between the two or more states, e.g., upon exposure to the proper stimuli. For example, a first stimuli (e.g., a first wavelength of light) may be used to activate the switchable entity, while a second stimuli (e.g., a second wavelength of light) may be used to deactivate the switchable entity, for instance, to a non-emitting state. Any suitable method may be used to activate the entity. For example, in one embodiment, incident light of a suitable wavelength may be used to activate the entity to emit light, i.e., the entity is "photoswitchable." Thus, the photoswitchable entity can be switched between different light-emitting or non-emitting states by incident light, e.g., of different wavelengths. The light may be monochromatic (e.g., produced using a laser) or polychromatic. In another embodiment, the entity may be activated upon stimulation by electric field and/or magnetic field. In other embodiments, the entity may be activated upon exposure to a suitable chemical environment, e.g., by adjusting the pH, or inducing a reversible chemical reaction involving the entity, etc. Similarly, any suitable method may be used to deactivate the entity, and the methods of activating and deactivating the entity need not be the same. For instance, the entity may be deactivated upon exposure to incident light of a suitable wavelength, or the entity may be deactivated by waiting a sufficient time.

Typically, a "switchable" entity can be identified by one of ordinary skill in the art by determining conditions under which an entity in a first state can emit light when exposed to an excitation wavelength, switching the entity from the first state to the second state, e.g., upon exposure to light of a switching wavelength, then showing that the entity, while in the second state can no longer emit light (or emits light at a much reduced intensity) when exposed to the excitation wavelength.

In one set of embodiments, as discussed, a switchable entity may be switched upon exposure to light. In some cases, the light used to activate the switchable entity may come from an external source, e.g., a light source such as a laser light source, another light-emitting entity proximate the switchable entity, etc. The second, light emitting entity, in some cases, may be a fluorescent entity, and in certain embodiments, the second, light-emitting entity may itself also be a switchable entity.

In some embodiments, the switchable entity includes a first, light-emitting portion (e.g., a fluorophore), and a second portion that activates or "switches" the first portion. For example, upon exposure to light, the second portion of the switchable entity may activate the first portion, causing the first portion to emit light. Examples of activator portions include, but are not limited to, Alexa Fluor 405 (Invitrogen), Alexa Fluor 488 (Invitrogen), Cy2 (GE Healthcare), Cy3 (GE Healthcare), Cy3B (GE Healthcare), Cy3.5 (GE Healthcare), or other suitable dyes. Examples of light-emitting portions include, but are not limited to, Cy5, Cy5.5 (GE Healthcare), Cy7 (GE Healthcare), Alexa Fluor 647 (Invitrogen), Alexa Fluor 680 (Invitrogen), Alexa Fluor 700 (Invitrogen), Alexa Fluor 750 (Invitrogen), Alexa Fluor 790 (Invitrogen), DiD, DiR, YOYO-3 (Invitrogen), YO-PRO-3 (Invitrogen), TOT-3 (Invitrogen), TO-PRO-3 (Invitrogen) or other suitable dyes. These may linked together, e.g., covalently, for example, directly, or through a linker, e.g., forming compounds such as, but not limited to, Cy5-Alexa Fluor 405, Cy5-Alexa Fluor 488, Cy5-Cy2, Cy5-Cy3, Cy5-Cy3.5, Cy5.5-Alexa Fluor 405, Cy5.5-Alexa Fluor 488, Cy5.5-Cy2, Cy5.5-Cy3, Cy5.5-Cy3.5, Cy7-Alexa Fluor 405, Cy7-Alexa Fluor 488, Cy7-Cy2, Cy7-Cy3, Cy7-Cy3.5, Alexa Fluor 647-Alexa Fluor 405, Alexa Fluor 647-Alexa Fluor 488, Alexa Fluor 647-Cy2, Alexa Fluor 647-Cy3, Alexa Fluor 647-Cy3.5, Alexa Fluor 750-Alexa Fluor 405, Alexa Fluor 750-Alexa Fluor 488, Alexa Fluor 750-Cy2, Alexa Fluor 750-Cy3, or Alexa Fluor 750-Cy3.5. Those of ordinary skill in the art will be aware of the structures of these and other compounds, many of which are available commercially. The portions may be linked via a covalent bond, or by a linker, such as those described in detail below. Other light-emitting or activator portions may include portions having two quaternized nitrogen atoms joined by a polymethine chain, where each nitrogen is independently part of a heteroaromatic moiety, such as pyrrole, imidazole, thiazole, pyridine, quinoine, indole, benzothiazole, etc., or part of a nonaromatic amine. In some cases, there may be 5, 6, 7, 8, 9, or more carbon atoms between the two nitrogen atoms.

In certain cases, the light-emitting portion and the activator portions, when isolated from each other, may each be fluorophores, i.e., entities that can emit light of a certain,

emission wavelength when exposed to a stimulus, for example, an excitation wavelength. However, when a switchable entity is formed that comprises the first fluorophore and the second fluorophore, the first fluorophore forms a first, light-emitting portion and the second fluorophore forms an activator portion that switches that activates or "switches" the first portion in response to a stimulus. For example, the switchable entity may comprise a first fluorophore directly bonded to the second fluorophore, or the first and second entity may be connected via a linker or a common entity. Whether a pair of light-emitting portion and activator portion produces a suitable switchable entity can be tested by methods known to those of ordinary skills in the art. For example, light of various wavelength can be used to stimulate the pair and emission light from the light-emitting portion can be measured to determined wither the pair makes a suitable switch.

As a non-limiting example, Cy3 and Cy5 may be linked together to form such an entity. In this example, Cy3 is an activator portion that is able to activate Cy5, the light-emission portion. Thus, light at or near the absorption maximum (e.g., near 532 nm light for Cy3) of the activation or second portion of the entity may cause that portion to activate the first, light-emitting portion, thereby causing the first portion to emit light (e.g., near 647 nm for Cy5). See, e.g., U.S. Pat. No. 7,838,302, incorporated herein by reference in its entirety. In some cases, the first, light-emitting portion can subsequently be deactivated by any suitable technique (e.g., by directing 647 nm red light to the Cy5 portion of the molecule).

Other non-limiting examples of potentially suitable activator portions include 1,5 IAEDANS, 1,8-ANS, 4-Methylumbelliferone, 5-carboxy-2,7-dichlorofluorescein, 5-Carboxyfluorescein (5-FAM), 5-Carboxynapthofluorescein, 5-Carboxytetramethylrhodamine (5-TAMRA), 5-FAM (5-Carboxyfluorescein), 5-HAT (Hydroxy Tryptamine), 5-Hydroxy Tryptamine (HAT), 5-ROX (carboxy-X-rhodamine), 5-TAMRA (5-Carboxytetramethylrhodamine), 6-Carboxyrhodamine 6G, 6-CR 6G, 6-JOE, 7-Amino-4-methylcoumarin, 7-Aminoactinomycin D (7-AAD), 7-Hydroxy-4-methylcoumarin, 9-Amino-6-chloro-2-methoxyacridine, ABQ, Acid Fuchsin, ACMA (9-Amino-6-chloro-2-methoxyacridine), Acridine Orange, Acridine Red, Acridine Yellow, Acriflavin, Acriflavin Feulgen SITSA, Alexa Fluor 350, Alexa Fluor 405, Alexa Fluor 430, Alexa Fluor 488, Alexa Fluor 500, Alexa Fluor 514, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 555, Alexa Fluor 568, Alexa Fluor 594, Alexa Fluor 610, Alexa Fluor 633, Alexa Fluor 635, Alizarin Complexon, Alizarin Red, AMC, AMCA-S, AMCA (Aminomethylcoumarin), AMCA-X, Aminoactinomycin D, Aminocoumarin, Aminomethylcoumarin (AMCA), Anilin Blue, Anthrocyl stearate, APTRA-BTC, APTS, Astrazon Brilliant Red 4G, Astrazon Orange R,

Astrazon Red 6B, Astrazon Yellow 7 GLL, Atabrine, ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 520, ATTO 532, ATTO 550, ATTO 565, ATTO 590, ATTO 594, ATTO 610, ATTO 611X, ATTO 620, ATTO 633, ATTO 635, ATTO 647, ATTO 647N, ATTO 655, ATTO 680, ATTO 700, ATTO 725, ATTO 740, ATTO-TAG CBQCA, ATTO-TAG FQ, Auramine, Aurophosphine G, Aurophosphine, BAO 9 (Bisaminophenyloxadiazole), BCECF (high pH), BCECF (low pH), Berberine Sulphate, Bimane, Bisbenzamide, Bisbenzimide (Hoechst), bis-BTC, Blancophor FFG, Blancophor SV, BOBO -1, BOBO -3, Bodipy 492/515, Bodipy 493/503, Bodipy 500/510, Bodipy 505/515, Bodipy 530/550, Bodipy 542/563, Bodipy 558/568, Bodipy 564/570, Bodipy 576/589, Bodipy 581/591, Bodipy 630/650-X, Bodipy 650/665-X, Bodipy 665/676, Bodipy Fl, Bodipy FL ATP, Bodipy Fl-Ceramide, Bodipy R6G, Bodipy TMR, Bodipy TMR-X conjugate, Bodipy TMR-X, SE, Bodipy TR, Bodipy TR ATP, Bodipy TR-X SE, BO-PRO -1, BO-PRO -3, Brilliant Sulphoflavin FF, BTC, BTC-5N, Calcein, Calcein Blue, Calcium Crimson, Calcium Green, Calcium Green-1 Ca^{2+} Dye, Calcium Green-2 Ca^{2+}, Calcium Green-5N Ca^{2+}, Calcium Green-C18 Ca^{2+}, Calcium Orange, Calcofluor White, Carboxy-X-rhodamine (5-ROX), Cascade Blue, Cascade Yellow, Catecholamine, CCF2 (GeneBlazer), CFDA, Chromomycin A, Chromomycin A, CL-NERF, CMFDA, Coumarin Phalloidin, CPM Methylcoumarin, CTC, CTC Formazan, Cy2, Cy3.1 8, Cy3.5, Cy3, Cy5.1 8, cyclic AMP Fluorosensor (FiCRhR), Dabcyl, Dansyl, Dansyl Amine, Dansyl Cadaverine, Dansyl Chloride, Dansyl DHPE, Dansyl fluoride, DAPI, Dapoxyl, Dapoxyl 2, Dapoxyl 3' DCFDA, DCFH (Dichlorodihydrofluorescein Diacetate), DDAO, DHR (Dihydorhodamine 123), Di-4-ANEPPS, Di-8-ANEPPS (non-ratio), DiA (4-Di-16-ASP), Dichlorodihydrofluorescein Diacetate (DCFH), DiD - Lipophilic Tracer, DiD (DiIC18(5)), DIDS, Dihydorhodamine 123 (DHR), DiI (DiIC18(3)), Dinitrophenol, DiO (DiOC18(3)), DiR, DiR (DiIC18(7)), DM-NERF (high pH), DNP, Dopamine, DTAF, DY-630-NHS, DY-635-NHS, DyLight 405, DyLight 488, DyLight 549, DyLight 633, DyLight 649, DyLight 680, DyLight 800, ELF 97, Eosin, Erythrosin, Erythrosin ITC, Ethidium Bromide, Ethidium homodimer -1 (EthD-1), Euchrysin, EukoLight, Europium (III) chloride, Fast Blue, FDA, Feulgen (Pararosaniline), FIF (Formaldehyd Induced Fluorescence), FITC, Flazo Orange, Fluo-3, Fluo-4, Fluorescein (FITC), Fluorescein Diacetate, Fluoro-Emerald, Fluoro-Gold (Hydroxystilbamidine), Fluor-Ruby, FluorX, FM 1-43, FM 4-46, Fura Red (high pH), Fura Red/Fluo-3, Fura-2, Fura-2/BCECF, Genacryl Brilliant Red B, Genacryl Brilliant Yellow 10GF, Genacryl Pink 3G, Genacryl Yellow 5GF, GeneBlazer (CCF2), Gloxalic Acid, Granular blue, Haematoporphyrin, Hoechst 33258, Hoechst 33342, Hoechst 34580, HPTS,

Hydroxycoumarin, Hydroxystilbamidine (FluoroGold), Hydroxytryptamine, Indo-1, high calcium, Indo-1, low calcium, Indodicarbocyanine (DiD), Indotricarbocyanine (DiR), Intrawhite Cf, JC-1, JO-JO-1, JO-PRO-1, LaserPro, Laurodan, LDS 751 (DNA), LDS 751 (RNA), Leucophor PAF, Leucophor SF, Leucophor WS, Lissamine Rhodamine, Lissamine Rhodamine B, Calcein/Ethidium homodimer, LOLO-1, LO-PRO-1, Lucifer Yellow, Lyso Tracker Blue, Lyso Tracker Blue-White, Lyso Tracker Green, Lyso Tracker Red, Lyso Tracker Yellow, LysoSensor Blue, LysoSensor Green, LysoSensor Yellow/Blue, Mag Green, Magdala Red (Phloxin B), Mag-Fura Red, Mag-Fura-2, Mag-Fura-5, Mag-Indo-1, Magnesium Green, Magnesium Orange, Malachite Green, Marina Blue, Maxilon Brilliant Flavin 10 GFF, Maxilon Brilliant Flavin 8 GFF, Merocyanin, Methoxycoumarin, Mitotracker Green FM, Mitotracker Orange, Mitotracker Red, Mitramycin, Monobromobimane, Monobromobimane (mBBr-GSH), Monochlorobimane, MPS (Methyl Green Pyronine Stilbene), NBD, NBD Amine, Nile Red, Nitrobenzoxadidole, Noradrenaline, Nuclear Fast Red, Nuclear Yellow, Nylosan Brilliant Iavin E8G, Oregon Green, Oregon Green 488-X, Oregon Green, Oregon Green 488, Oregon Green 500, Oregon Green 514, Pacific Blue, Pararosaniline (Feulgen), PBFI, Phloxin B (Magdala Red), Phorwite AR, Phorwite BKL, Phorwite Rev, Phorwite RPA, Phosphine 3R, PKH26 (Sigma), PKH67, PMIA, Pontochrome Blue Black, POPO-1, POPO-3, PO-PRO-1, PO-PRO-3, Primuline, Procion Yellow, Propidium Iodid (PI), PyMPO, Pyrene, Pyronine, Pyronine B, Pyrozal Brilliant Flavin 7GF, QSY 7, Quinacrine Mustard, Resorufin, RH 414, Rhod-2, Rhodamine, Rhodamine 110, Rhodamine 123, Rhodamine 5 GLD, Rhodamine 6G, Rhodamine B, Rhodamine B 200, Rhodamine B extra, Rhodamine BB, Rhodamine BG, Rhodamine Green, Rhodamine Phallicidine, Rhodamine Phalloidine, Rhodamine Red, Rhodamine WT, Rose Bengal, S65A, S65C, S65L, S65T, SBFI, Serotonin, Sevron Brilliant Red 2B, Sevron Brilliant Red 4G, Sevron Brilliant Red B, Sevron Orange, Sevron Yellow L, SITS, SITS (Primuline), SITS (Stilbene Isothiosulphonic Acid), SNAFL calcein, SNAFL-1, SNAFL-2, SNARF calcein, SNARF1, Sodium Green, SpectrumAqua, SpectrumGreen, SpectrumOrange, Spectrum Red, SPQ (6-methoxy-N-(3-sulfopropyl)quinolinium), Stilbene, Sulphorhodamine B can C, Sulphorhodamine Extra, SYTO 11, SYTO 12, SYTO 13, SYTO 14, SYTO 15, SYTO 16, SYTO 17, SYTO 18, SYTO 20, SYTO 21, SYTO 22, SYTO 23, SYTO 24, SYTO 25, SYTO 40, SYTO 41, SYTO 42, SYTO 43, SYTO 44, SYTO 45, SYTO 59, SYTO 60, SYTO 61, SYTO 62, SYTO 63, SYTO 64, SYTO 80, SYTO 81, SYTO 82, SYTO 83, SYTO 84, SYTO 85, SYTOX Blue, SYTOX Green, SYTOX Orange, Tetracycline, Tetramethylrhodamine (TAMRA), Texas Red, Texas Red-X conjugate, Thiadicarbocyanine (DiSC3), Thiazine Red

R, Thiazole Orange, Thioflavin 5, Thioflavin S, Thioflavin TCN, Thiolyte, Thiozole Orange, Tinopol CBS (Calcofluor White), TMR, TO-PRO-1, TO-PRO-3, TO-PRO-5, TOTO-1, TOTO-3, TRITC (tetramethylrodamine isothiocyanate), True Blue, TruRed, Ultralite, Uranine B, Uvitex SFC, WW 781, X-Rhodamine, XRITC, Xylene Orange, Y66F, Y66H, Y66W, YO-PRO-1, YO-PRO-3, YOYO-1, YOYO-3, SYBR Green, Thiazole orange (interchelating dyes), or combinations thereof.

Another aspect of the invention is directed to a computer-implemented method. For instance, a computer and/or an automated system may be provided that is able to automatically and/or repetitively perform any of the methods described herein. As used herein, "automated" devices refer to devices that are able to operate without human direction, i.e., an automated device can perform a function during a period of time after any human has finished taking any action to promote the function, e.g. by entering instructions into a computer to start the process. Typically, automated equipment can perform repetitive functions after this point in time. The processing steps may also be recorded onto a machine-readable medium in some cases.

For example, in some cases, a computer may be used to control imaging of the sample, e.g., using fluorescence microscopy, STORM or other super-resolution techniques such as those described herein. In some cases, the computer may also control operations such as drift correction, physical registration, hybridization and cluster alignment in image analysis, cluster decoding (e.g., fluorescent cluster decoding), error detection or correction (e.g., as discussed herein), noise reduction, identification of foreground features from background features (such as noise or debris in images), or the like. As an example, the computer may be used to control activation and/or excitation of signaling entities within the sample, and/or the acquisition of images of the signaling entities. In one set of embodiments, a sample may be excited using light having various wavelengths and/or intensities, and the sequence of the wavelengths of light used to excite the sample may be correlated, using a computer, to the images acquired of the sample containing the signaling entities. For instance, the computer may apply light having various wavelengths and/or intensities to a sample to yield different average numbers of signaling entities in each region of interest (e.g., one activated entity per location, two activated entities per location, etc.). In some cases, this information may be used to construct an image and/or determine the locations of the signaling entities, in some cases at high resolutions, as noted above.

In some aspects, the sample is positioned on a microscope. In some cases, the microscope may contain one or more channels, such as microfluidic channels, to direct or

control fluid to or from the sample. For instance, in one embodiment, nucleic acid probes such as those discussed herein may be introduced and/or removed from the sample by flowing fluid through one or more channels to or from the sample. In some cases, there may also be one or more chambers or reservoirs for holding fluid, e.g., in fluidic communication with the channel, and/or with the sample. Those of ordinary skill in the art will be familiar with channels, including microfluidic channels, for moving fluid to or from a sample.

As used herein, "microfluidic," "microscopic," "microscale," the "micro-" prefix (for example, as in "microchannel"), and the like generally refers to elements or articles having widths or diameters of less than about 1 mm, and less than about 100 microns (micrometers) in some cases. In some embodiments, larger channels may be used instead of, or in conjunction with, microfluidic channels for any of the embodiments discussed herein. For examples, channels having widths or diameters of less than about 10 mm, less than about 9 mm, less than about 8 mm, less than about 7 mm, less than about 6 mm, less than about 5 mm, less than about 4 mm, less than about 3 mm, or less than about 2 mm may be used in certain instances. In some cases, the element or article includes a channel through which a fluid can flow. In all embodiments, specified widths can be a smallest width (i.e. a width as specified where, at that location, the article can have a larger width in a different dimension), or a largest width (i.e. where, at that location, the article has a width that is no wider than as specified, but can have a length that is greater). Thus, for instance, the microfluidic channel may have an average cross-sectional dimension (e.g., perpendicular to the direction of flow of fluid in the microfluidic channel) of less than about 1 mm, less than about 500 microns, less than about 300 microns, or less than about 100 microns. In some cases, the microfluidic channel may have an average diameter of less than about 60 microns, less than about 50 microns, less than about 40 microns, less than about 30 microns, less than about 25 microns, less than about 10 microns, less than about 5 microns, less than about 3 microns, or less than about 1 micron.

A "channel," as used herein, means a feature on or in an article (e.g., a substrate) that at least partially directs the flow of a fluid. In some cases, the channel may be formed, at least in part, by a single component, e.g. an etched substrate or molded unit. The channel can have any cross-sectional shape, for example, circular, oval, triangular, irregular, square or rectangular (having any aspect ratio), or the like, and can be covered or uncovered (i.e., open to the external environment surrounding the channel). In embodiments where the channel is completely covered, at least one portion of the channel can have a cross-section that is

completely enclosed, and/or the entire channel may be completely enclosed along its entire length with the exception of its inlet and outlet.

A channel may have any aspect ratio, e.g., an aspect ratio (length to average cross-sectional dimension) of at least about 2:1, more typically at least about 3:1, at least about 5:1, at least about 10:1, etc. As used herein, a "cross-sectional dimension," in reference to a fluidic or microfluidic channel, is measured in a direction generally perpendicular to fluid flow within the channel. A channel generally will include characteristics that facilitate control over fluid transport, e.g., structural characteristics and/or physical or chemical characteristics (hydrophobicity vs. hydrophilicity) and/or other characteristics that can exert a force (e.g., a containing force) on a fluid. The fluid within the channel may partially or completely fill the channel. In some cases the fluid may be held or confined within the channel or a portion of the channel in some fashion, for example, using surface tension (e.g., such that the fluid is held within the channel within a meniscus, such as a concave or convex meniscus). In an article or substrate, some (or all) of the channels may be of a particular size or less, for example, having a largest dimension perpendicular to fluid flow of less than about 5 mm, less than about 2 mm, less than about 1 mm, less than about 500 microns, less than about 200 microns, less than about 100 microns, less than about 60 microns, less than about 50 microns, less than about 40 microns, less than about 30 microns, less than about 25 microns, less than about 10 microns, less than about 3 microns, less than about 1 micron, less than about 300 nm, less than about 100 nm, less than about 30 nm, or less than about 10 nm or less in some cases. In one embodiment, the channel is a capillary.

A variety of materials and methods, according to certain aspects of the invention, can be used to form devices or components containing microfluidic channels, chambers, etc. For example, various devices or components can be formed from solid materials, in which the channels can be formed via micromachining, film deposition processes such as spin coating and chemical vapor deposition, physical vapor deposition, laser fabrication, photolithographic techniques, etching methods including wet chemical or plasma processes, electrodeposition, and the like. See, for example, *Scientific American*, 248:44-55, 1983 (Angell, *et al*).

In one set of embodiments, various structures or components can be formed of a polymer, for example, an elastomeric polymer such as polydimethylsiloxane ("PDMS"), polytetrafluoroethylene ("PTFE" or Teflon®), or the like. For instance, according to one embodiment, a channel such as a microfluidic channel may be implemented by fabricating the fluidic system separately using PDMS or other soft lithography techniques (details of soft lithography techniques suitable for this embodiment are discussed in the references entitled

"Soft Lithography," by Younan Xia and George M. Whitesides, published in the *Annual Review of Material Science*, 1998, Vol. 28, pages 153-184, and "Soft Lithography in Biology and Biochemistry," by George M. Whitesides, Emanuele Ostuni, Shuichi Takayama, Xingyu Jiang and Donald E. Ingber, published in the *Annual Review of Biomedical Engineering*, 2001, Vol. 3, pages 335-373; each of these references is incorporated herein by reference).

Other examples of potentially suitable polymers include, but are not limited to, polyethylene terephthalate (PET), polyacrylate, polymethacrylate, polycarbonate, polystyrene, polyethylene, polypropylene, polyvinylchloride, cyclic olefin copolymer (COC), polytetrafluoroethylene, a fluorinated polymer, a silicone such as polydimethylsiloxane, polyvinylidene chloride, bis-benzocyclobutene ("BCB"), a polyimide, a fluorinated derivative of a polyimide, or the like. Combinations, copolymers, or blends involving polymers including those described above are also envisioned. The device may also be formed from composite materials, for example, a composite of a polymer and a semiconductor material.

In some embodiments, various microfluidic structures or components of the device are fabricated from polymeric and/or flexible and/or elastomeric materials, and can be conveniently formed of a hardenable fluid, facilitating fabrication via molding (e.g. replica molding, injection molding, cast molding, etc.). The hardenable fluid can be essentially any fluid that can be induced to solidify, or that spontaneously solidifies, into a solid capable of containing and/or transporting fluids contemplated for use in and with the fluidic network. In one embodiment, the hardenable fluid comprises a polymeric liquid or a liquid polymeric precursor (i.e. a "prepolymer"). Suitable polymeric liquids can include, for example, thermoplastic polymers, thermoset polymers, waxes, metals, or mixtures or composites thereof heated above their melting point. As another example, a suitable polymeric liquid may include a solution of one or more polymers in a suitable solvent, which solution forms a solid polymeric material upon removal of the solvent, for example, by evaporation. Such polymeric materials, which can be solidified from, for example, a melt state or by solvent evaporation, are well known to those of ordinary skill in the art. A variety of polymeric materials, many of which are elastomeric, are suitable, and are also suitable for forming molds or mold masters, for embodiments where one or both of the mold masters is composed of an elastomeric material. A non-limiting list of examples of such polymers includes polymers of the general classes of silicone polymers, epoxy polymers, and acrylate polymers. Epoxy polymers are characterized by the presence of a three-membered cyclic ether group commonly referred to as an epoxy group, 1,2-epoxide, or oxirane. For example, diglycidyl

ethers of bisphenol A can be used, in addition to compounds based on aromatic amine, triazine, and cycloaliphatic backbones. Another example includes the well-known Novolac polymers. Non-limiting examples of silicone elastomers suitable for use according to the invention include those formed from precursors including the chlorosilanes such as methylchlorosilanes, ethylchlorosilanes, phenylchlorosilanes, etc.

Silicone polymers are used in certain embodiments, for example, the silicone elastomer polydimethylsiloxane. Non-limiting examples of PDMS polymers include those sold under the trademark Sylgard by Dow Chemical Co., Midland, MI, and particularly Sylgard 182, Sylgard 184, and Sylgard 186. Silicone polymers including PDMS have several beneficial properties simplifying fabrication of various structures of the invention. For instance, such materials are inexpensive, readily available, and can be solidified from a prepolymeric liquid via curing with heat. For example, PDMSs are typically curable by exposure of the prepolymeric liquid to temperatures of about, for example, about 65 °C to about 75 °C for exposure times of, for example, at least about an hour. Also, silicone polymers, such as PDMS, can be elastomeric and thus may be useful for forming very small features with relatively high aspect ratios, necessary in certain embodiments of the invention. Flexible (e.g., elastomeric) molds or masters can be advantageous in this regard.

One advantage of forming structures such as microfluidic structures or channels from silicone polymers, such as PDMS, is the ability of such polymers to be oxidized, for example by exposure to an oxygen-containing plasma such as an air plasma, so that the oxidized structures contain, at their surface, chemical groups capable of cross-linking to other oxidized silicone polymer surfaces or to the oxidized surfaces of a variety of other polymeric and non-polymeric materials. Thus, structures can be fabricated and then oxidized and essentially irreversibly sealed to other silicone polymer surfaces, or to the surfaces of other substrates reactive with the oxidized silicone polymer surfaces, without the need for separate adhesives or other sealing means. In most cases, sealing can be completed simply by contacting an oxidized silicone surface to another surface without the need to apply auxiliary pressure to form the seal. That is, the pre-oxidized silicone surface acts as a contact adhesive against suitable mating surfaces. Specifically, in addition to being irreversibly sealable to itself, oxidized silicone such as oxidized PDMS can also be sealed irreversibly to a range of oxidized materials other than itself including, for example, glass, silicon, silicon oxide, quartz, silicon nitride, polyethylene, polystyrene, glassy carbon, and epoxy polymers, which have been oxidized in a similar fashion to the PDMS surface (for example, via exposure to an oxygen-containing plasma). Oxidation and sealing methods useful in the context of the

present invention, as well as overall molding techniques, are described in the art, for example, in an article entitled "Rapid Prototyping of Microfluidic Systems and Polydimethylsiloxane," *Anal. Chem.*, 70:474-480, 1998 (Duffy *et al.*), incorporated herein by reference.

The following documents are each incorporated herein by reference in their entireties: U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*; and Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*

In addition, incorporated herein by reference in their entireties are U.S. Provisional Patent Application Serial No. 62/031,062, filed July 30, 2014, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/050,636, filed September 15, 2014, entitled "Probe Library Construction," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/142,653, filed April 3, 2015, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; and a PCT application filed on even date herewith, entitled "Probe Library Construction," by Zhuang, *et al.*

The following examples are intended to illustrate certain embodiments of the present invention, but do not exemplify the full scope of the invention.

EXAMPLE 1

The example presents a platform to enable the simultaneous detection of the number and spatial organization of thousands of distinct mRNAs within single cells with high efficiency and low error-rate using a novel form of highly multiplexed fluorescence *in situ* hybridization (FISH). This example accomplishes these measurements by integrating and innovating methods for massively parallel probe synthesis, super-resolution imaging, and self-correcting error-checking codes.

Here, these examples present methods for the simultaneous detection of some or all of the thousands of unique RNAs expressed in a cell. This approach not only promises to revolutionize the throughput of the already effective single-molecule FISH (smFISH) approach, but also allows researchers to benefit from the hypothesis free discovery approach which has made other whole-genome systems approaches to biology so effective. For example, this whole genome approach may allow researchers to discover RNAs whose expression levels and/or subcellular localization patterns change under certain conditions of interest, such as disease states, without knowing, *a priori*, which mRNA will change in

abundance or localization. Simultaneous measurements of hundreds of genes within a single cell also allow for the identification of correlations between genes in expression and localization patterns in some cases.

This can be achieved using methods for highly multiplexed smFISH via the sequential hybridization of orthogonal detection probes and super-resolution imaging, reducing the cost of probe synthesis, and the development of a highly automated system to minimize demands on the user, as discussed herein. This provides an integrated platform to handle the bioinformatics of probe design, the mathematics of error-correcting codes, the complexity of image registration and analysis, and the cumbersome fluid handling through a simple suite of user friendly interfaces. This integration allows easy operation with limited user training and facilitates the rapid collection of data.

This example illustrates: (1) computational design of "codewords" to attach to all RNA targets in the cell that will allow unique identification of each RNA with some degree of experimental error tolerance, (2) translation of these codewords into nucleotide sequences and synthesis of the required single-stranded (ss) oligonucleotide (e.g. ssDNA) probes, (3) sample fixation and hybridization of these probes to the RNA targets *in situ*, (4) read-out of these codewords via successive rounds of hybridization of distinct fluorescent probes imaged with conventional fluorescence microscopy or super-resolution fluorescence microscopy, and (5) automated decoding of measured codewords combined with computational error correction to uniquely and robustly identify individual mRNAs.

In the first step, a "codeword" is assigned to every RNA that is to be labeled. In a typical design these may be strings of N binary letters or positions. Codewords may be chosen from the same wide range of existing error tolerant or error-correcting encoding schemes developed for digital storage and communication, e.g., using Hamming codes or the like. For example, actin-RNA may be assigned the binary codeword 11001010. Each codeword may be unique and separated from the other codewords by a Hamming distance, h, which measures the number of letters or positions that must be incorrectly read for one codeword to be misinterpreted as a different one. A Hamming distance greater than 1 between all codewords allows for some measurement errors to be detected—since simple errors would produce codewords that are not used to encode RNAs. For a Hamming distance larger than 2, it is also possible to correct some errors, as codewords with one error will be closest in Hamming distance to a single, unique codeword. The total number of different RNAs to be detected from the transcriptome and the amount of error correction desired

determines the length of the codewords. Information theory provides several efficient algorithms for assembling error-correcting binary codebooks.

In the second step, this encoding scheme is translated into a set of oligonucleotide (e.g. DNA) probe sequences, which can be called primary probes or encoding probes, each of which not only targets a probe to the RNA of interest but also encodes the unique binary codeword within a set of secondary binding sites (Fig. 1). For example, first designed may be primary binding sequences for each targeted mRNA. These sequences are "target sequences" that are comprised of complementary nucleotide sequences to their target RNAs computationally selected to satisfy a stringent set of hybridization conditions, including uniqueness in the target genome. To improve the efficiency of hybridization to individual mRNAs, multiple primary target sequences are designed for each individual RNA. Then each position within the set of codewords is assigned a unique oligonucleotide (e.g. DNA) sequence, which is called a read sequence. These tags are designed as to have no interaction with endogenous mRNA sequences or each other. For instance, for all the value "1"s in a codeword of an individual mRNA, the corresponding read sequence is attached to the primary targeting sequences against that mRNA. In general, each probe will contain a target sequence and one or more read sequences. If the total length of the necessary read sequences and the primary target sequence exceeds synthesis capabilities, then subsets of the read sequences can be appended to distinct target sequences. For example, consider the potential codeword 11001010 for actin. Probe sequences for this RNA could contain the read sequences corresponding to positions 1, 2, 5, and 7 in the codeword attached to a variety of target sequences specific to actin. After all the sequences have been designed, the resulting complex set of unique custom oligonucleotide (e.g. DNA) sequences is manufactured and amplified using methods as described below.

In the third step, the resulting pool of DNA is hybridized, e.g., to fixed, permeabilized cells. In this process, individual probes may be attached to every RNA in the cell by hybridization of their corresponding target sequences with the RNA while the read sequences remain free to bind the appropriate secondary probes as discussed below.

In the fourth step—the read-out step—fluorescently labeled secondary nucleic acid probes (also called readout probes) are successively hybridized to the read sequences attached to the target sequences that binds to the mRNA targets in the above step. When simultaneously imaging a large number of different RNA species in cells, the density of labeled RNAs may exceed that at which each RNA can be resolved via conventional imaging methods. Thus, this may be performed using a super-resolution imaging method, for example

STORM (stochastic optical reconstruction microscopy), to resolve the labeled molecules. After each round of hybridization and imaging with the secondary probes, the fluorophores are quenched or otherwise inactivated either via chemical or optical techniques such as oxidation, chemically bleaching, photobleaching, stringent washing or enzymatic digestion, etc. The sample is then stained with the next secondary probe, and the cycle continues until all positions of the codewords have been read out. In the simplest incarnation, there will be one hybridization step for each position within the codeword, e.g. 8 hybridization steps for an 8-letter codeword (Fig. 1).

Fig. 1 shows schematic diagrams of this example. Fig. 1A shows that every position of the codewords is assigned a unique oligonucleotide sequence when this position has a value "1." All mRNA codewords are then translated into combinations of read sequences, which are attached to the targeting sequence. Fig. 1B shows various steps of the labeling scheme of this example. In the first step, all mRNAs (I - III) are tagged with multiple oligonucleotide (e.g. ssDNA) probes comprising a primary targeting sequence, which hybridizes to the RNA of interest, and a "tail" (i.e. containing read sequences) carrying the translated codeword, which does not interact with endogenous nucleotide sequences. In the next step, the first secondary probe is added, which can bind all probes whose tails have a read sequence corresponding to the value of "1" in the first position. The dyes on these secondary probes are imaged and bleached, then the next secondary probe is added to bind probes attached to mRNA which have a value of "1" in the second position of their assigned codeword, and so on.

In the final step, the microscopy images from each staining and imaging round are aligned, for example, computationally (e.g. using fiducial beads or other markers tracked during image acquisition), and the clusters of localizations resolved by conventional fluorescence microscopy or super-resolution imaging (e.g. STORM) from the different rounds are identified. These clusters of localizations arise from individual target mRNA molecules, and the hybridization rounds in which a spot was detected in a given cluster correspond to the "1" in the codeword for that mRNA. If there are no missed-detection events or false positive signal in the images, this codeword will perfectly match one of the expected codewords. Fig. 1 describes an example in which the codeword has three letters, i.e. three positions, and the three target mRNAs have codewords 110, 101, and 011 assigned to them. In real experimental examples, the codeword could contain more digits. For example, the mRNA for actin can be assigned the codeword, 11001010. In that case, detected clusters containing overlapping localization signals in the 1st, 2nd, 5th and 7th hybridization steps

(meaning the 1st, 2nd, 5th and 7th secondary probes bound to this site) can be identified as individual actin mRNA molecules, since the pattern of positive bindings matches the codewode of actin (11001010). In addition, if there are missed-detection events or false positive signals in the image data, these aberrations can be corrected by the implemented error-correction scheme. For example, clusters of localizations with a detected codeword that has only one digit discrepancy from 11001010 (such as 11000010 or 11101010) can also be identified as actin mRNA since all other valid codeword in this example differ from the detected pattern in two more positions.

EXAMPLE 2

This example describes another alternative approach that differs in several of the steps described above. This approach begins with the first step, construction of the codewords to the desired mRNA targets, as described above.

In the second step of this approach, nucleic acid probes are designed that bind uniquely to the mRNA targets of interest, as described above. However, instead of appending unique read sequences to these targeting sequences, unique pools or groups of probes are constructed from these target sequences. Each pool comprises all or a subset of the sequences that target all mRNAs which contain the same value at a given position in their codeword. For example, the first pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the first position of their codewords, e.g. 110 and 101 but not 011; the second pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the second position of their codewords, e.g. 110 and 011 but not 101; the third pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the third position of their codewords, e.g. 011 and 101 but not 110 (Fig. 1C). As another example, consider the potential codeword 11001010 for actin. Probes that target this mRNA would be included in pools 1, 2, 5, and 7 but not in pools 3, 4, 6, and 8. The same target to a given mRNA may or may not be included in pools. For example, a probe that targets the same region of actin may be included in pools 1, 2, 5, and 7 or any subset of these pools. After all of the pools have been designed, each complex set of unique, custom oligonucleotide sequences is manufactured and amplified using methods as described below.

In the third step of this approach, the first pool of probes is hybridized, e.g., to fixed, permeabilized cells. In this process, the fluorophores attached to each of the probes in this pool are bound to each of the targets of that pool. The binding of these probes is then determined by fluorescence microscopy. As described above, these images can be collected either via a range of methods including both conventional fluorescence imaging or super-

resolution imaging methods such as STORM. After a round of imaging, the probes from the first pool are inactivated or removed from the sample via the methods described above. This process is then repeated for each successive pool of probes until some or all of the pools have been applied to the sample and imaged such that all positions in the codewords have been read out. In the simplest incarnation, there will be one hybridization and imaging step for each position in the codeword, e.g. 3 rounds of hybridization and imaging for a codeword with 3 positions (Fig 1C) or 8 rounds of hybridization and imaging for a codeword with 8 positions.

The final step of this approach is identical to that described above.

EXAMPLE 3

In this example, 14 genes (PGK1, H3F3B, PKM, ENO1, GPI, EEF2, GNAS, HSPA8, GAPDH, CALM1, RHOA, PPIA, UBA52, and VCP) were encoded using a subset of the (8,4) SECDED code (Figs. 2A-2E). To determine the accuracy of these measurements, the measured abundances of these 14 mRNAs were compared to the abundances measured from bulk RNA-seq of A549 cells (published ENCODE data). Remarkably, it was found that there was excellent agreement between these two measurements, as the transcripts count measured using the sequential hybridization approach correlated with gene expression measured using RNA-seq with a Pearson correlation coefficient r of 0.75 (Fig. 2F). Gene expression from 3 other cells were also measured, and it was found that the gene expression of these 14 genes was highly correlated among the cells with a r of 0.96 (Fig. 2G).

Codebook Design. Each mRNA in the target set was assigned a binary codeword using a Single Error Correction Double Error Detection (SECDED) code. SECDED is an extended Hamming codebook with an additional parity bit. Briefly, Matlab's Communications System toolbox was used to generate SECDED codes of either 8 or 16 letters or positions. In both cases, only those codewords containing four 1s were used. These words were assigned at random to mRNAs in the target set. [0 1 0 1 1 1 0 0] is an example of the 8-letter codewords used (i.e., these codewords each contained four 1s and four 0s.) [0 1 0 1 1 1 0 0 0 0 0 0 0 0 0 0] is an example of the 16-letter codewords used (i.e., each codeword contained four 1s and twelve 0s). Not every codeword was necessarily assigned to an mRNA.

Computational Assembly of ssDNA Primary Probe Sequence. The number of primary nucleic acid probes used for hybridization with mRNA targets ranged from 200 to 2000 unique oligonucelotides, depending on the experiment. For example, to label 14 mRNAs with 28 oligos targeting each gene, 392 unique sequences were used. Large number

of oligos with unique sequences were purchased in a pool from LC Sciences or CustomArray. However, array synthesized oligos were in minute quantity that was insufficient for *in situ* hybridization. The protocol for their amplification are described below.

Each primary probe contained three components: flanking primer sequences to allow enzymatic amplification of probes, targeting sequence for *in situ* hybridization to mRNAs, and secondary tag sequence containing one or more read sequences for sequential readout of codewords.

The following is an example of a primary probe:

GTTGGCGACGAAAGCACTGCGATTGGAACCGTCCCAAGCGTTGCG
CTTAATGGATCATCAATTTTGTCTCACTACGACGGTCAATCGCGCTGCATACTTG
CGTCGGTCGGACAAACGAGG

(SEQ ID NO: 1)

The components are arranged in the following order: forward primer (not underlined), secondary read sequence 1 (underlined), mRNA targeting sequence (not underlined), secondary read sequence 2 (underlined), and reverse primer (not underlined). The secondary read sequences are the reverse complement of the corresponding secondary probes. Since only codewords that contained four '1's were used, the primary probes for each mRNA needed to contain 4 different secondary read sequences in this example. However, in order to reduce the overall length of the primary probes, the pool of targeting sequences for each mRNA target was split at random into two pools. Two secondary read sequences are attached each probe in one of the two pools and the other two secondary read sequences are attached the probes in the other pool. The design criteria for each component are described below.

Primer Design. Specific index primers were generated by a collection of 240,000 published sequences of orthogonal 25-bp long sequences. These sequences were trimmed to 20 bp, selected for a narrow 70 to 80 °C melting temperature, the absence of consecutive repeats of 3 or more base, and the presence of a GC clamp, i.e. one of the two 3' terminal bases must be G or C. To further improve specificity, these sequences were then screened against the human genome using BLAST+ (Camacho et al 2009), and primers with 14 or more contiguous bases of homology were eliminated. In a subsequent screening via BLAST+, primers that shared 11 or more contiguous bases or more than 5 bases at the 3' end of any other primer or the T7 promoter were also removed.

Secondary Probes Design. 30-bp long secondary probe sequences were created by concatenating fragments of the orthogonal primer set described above. These secondaries were then screened for orthogonality with other secondaries (no more than 11 basepairs of homology) and for potential off-target binding sites in the human genome (no more than 14 basepairs of homology). Secondary sequences used in this example are provided in Table 1.

Table 1

Bit	Secondary sequences	Sequence number
B1	CGCAACGCTTGGGACGGTTCCAATCGGATC	SEQ ID NO: 2
B2	CGAATGCTCTGGCCTCGAACGAACGATAGC	SEQ ID NO: 3
B3	ACAAATCCGACCAGATCGGACGATCATGGG	SEQ ID NO: 4
B4	CAAGTATGCAGCGCGATTGACCGTCTCGTT	SEQ ID NO: 5
B5	TGCGTCGTCTGGCTAGCACGGCACGCAAAT	SEQ ID NO: 6
B6	AAGTCGTACGCCGATGCGCAGCAATTCACT	SEQ ID NO: 7
B7	CGAAACATCGGCCACGGTCCCGTTGAACTT	SEQ ID NO: 8
B8	ACGAATCCACCGTCCAGCGCGTCAAACAGA	SEQ ID NO: 9
B9	CGCGAAATCCCCGTAACGAGCGTCCCTTGC	SEQ ID NO: 10
B10	GCATGAGTTGCCTGGCGTTGCGACGACTAA	SEQ ID NO: 11
B11	CCGTCGTCTCCGGTCCACCGTTGCGCTTAC	SEQ ID NO: 12
B12	GGCCAATGGCCCAGGTCCGTCACGCAATTT	SEQ ID NO: 13
B13	TTGATCGAATCGGAGCGTAGCGGAATCTGC	SEQ ID NO: 14
B14	CGCGCGGATCCGCTTGTCGGGAACGGATAC	SEQ ID NO: 15
B15	GCCTCGATTACGACGGATGTAATTCGGCCG	SEQ ID NO: 16
B16	GCCCGTATTCCCGCTTGCGAGTAGGGCAAT	SEQ ID NO: 17

mRNA Targeting Sequence Design. To determine the relative abundance of all the isoforms of all genes expressed in these cell lines, transcriptome profiling data from the ENCODE project for total RNA from A549 and IMR90 cells was processed using the publicly available software *cufflinks*, along with human genome annotations from gencode v18. Gene models corresponding to the most highly expressed isoform were used to build a sequence library in FASTA format recording the dominant isoform of every gene. Genes of interest were selected from this library. These genes were partitioned into 1kb segments, then

the software *OligoArray2.1* was used to generate primary probe sequences for the human transcriptome with the following constraints: 30-bp or 40-bp length, depending on the experiment; probe-target melting temperatures greater than 70 °C (variable parameter); no cross hybridization targets with melting temperatures greater than 72 °C (variable parameter); no predicted internal secondary structures with melting temperatures greater than 76 °C (variable parameter); and no single-nucleotide contiguous repeats of 6 or more bases. After OligoArray probe selection, all potential probes that mapped to a different gene were rejected while all potential probes with multiple alignments to the same gene were retained. A BLAST database was assembled from the FASTA library of all expressed genes to screen for probes' uniqueness. For each gene, 14 to 28 targeting sequences produced during the OligoArray processing were selected.

Probe Synthesis – index PCR. The template for specific probe sets were selected from the complex oligopool via limited-cycle PCR. Briefly, 0.5 to1 ng of the complex oligopool was combined with 0.5 micromolar of each primer. The forward primer matched the priming sequence for the desired subset while the reverse primer was a 5' concatenation of this sequence with a T7 promoter. To avoid the generation of G-quadruplets, which can be difficult to synthesize, the terminal Gs required in the T7 promoter were generated from Gs located at the 5' of the priming region where appropriate. All primers were synthesized by IDT. A 50 microliter reaction volume was amplified either using the KAPA real-time library amplification kit (KAPA Biosystems; KK2701) or via a homemade qPCR mix which included 0.8X EvaGreen (Biotum; 31000-T) and the hot-start Phusion polymerase (New England Biolabs; M0535S). Amplification was followed in real time using Agilent's MX300P or Biorad's CFX Connect. Individual samples were removed immediately before the plateau in amplification to minimize distortion of template abundance due to over-amplification. Individual templates were purified with columns according to the manufacturer's instructions (Zymo DNA Clean and Concentrator; D4003) and eluted in RNase-free deionized water.

Amplification via *in-vitro* transcription. The template was then amplified via *in vitro* transcription. Briefly, 0.5 to 1 microgram of template DNA was amplified into 100-200 microgram of RNA in a single 20-30 microliter reaction with a high yield RNA polymerase (New England Biolabs; E2040S). Reactions were supplemented with 1X RNase inhibitor (Promega RNasin; N2611). Amplification was typically run for 4 to 16 hours at 37 °C to maximize the yield. The RNA was not purified after the reaction and was either stored at -80 °C or immediately converted into DNA as described below.

Reverse Transcription. 1 to 2 nmol of fluorescently-labeled ssDNA probe was created from the above *in vitro* transcription reactions using the reverse transcriptase Maxima H- (Thermo Scientific; EP0751). This enzyme was used because of its higher processivity and temperature resistance, which allowed the conversion of large quantities of RNA into DNA within small volumes at temperatures that disfavor secondary structure formation. The unpurified RNA created above was supplemented with 1.6 mM of each dNTP, 1-2 nmol of fluorescently labeled forward primer, 300 units of Maxima H-, 60 units of RNasin, and a final 1X concentration of the Maxima RT buffer. The final 75 microliter volume was incubated at 50 °C for 60 minutes.

Strand Selection and Purification. The template RNA in the reaction above was then removed from the DNA via alkaline hydrolysis. 75 microliters of 0.25 M EDTA and 0.5 N NaOH were added to each reverse transcription reaction, and the sample was incubated at 95 °C for 10 minutes. The reaction was immediately neutralized by purifying the ssDNA probe with a modified version of the Zymo Oligo Clean and Concentrator protocol. Specifically, the 5-microgram capacity column was replaced with a 25-microgram or 100-microgram capacity DNA column as appropriate. The remainder of the protocol was run according to the manufacturer's instructions. Probe was eluted in 100 microliter RNase-free deionized water and evaporated in a vacuum concentrator. The final pellet was resuspended in 10 microliters of RNase-free water and stored at -20 °C. Denaturing poly-acrylimide gel electrophoresis and absorption spectroscopy revealed that this protocol typically produced 90-100% incorporation of the fluorescent primer into full length probe and 75-90% recovery of the total fluorescent probe. Thus, without exceeding a 150-microliter reaction volume, this protocol can be used to create ~2 nmol of fluorescent probe.

Cell culture and fixation. A549 and IMR90 cells (American Type Culture Collection) were cultured with Dulbecco's Modified Eagle Medium and Eagle's Minimum Essential Medium respectively. Cells were incubated at 37 °C with 5% CO_2 for 36-48 hours. Cells were fixed in 3% paraformaldehyde (Electron Microscopy Sciences) in PBS for 15 minutes, washed with PBS, and permeabilized in 70% ethanol overnight at 4 °C.

Fluorescence In Situ Hybridization (FISH) – primary (encoding) probes. Cells were hydrated in wash buffers (2xSSC, 50% formamide) for 10 minutes, labelled with primary oligos (0.5 nM per sequence) in hybridization buffers (2xSSC, 50% formamide, 1 mg/mL yeast tRNA, and 10% dextran sulfate) overnight at 37 °C, washed with wash buffers at 47 °C for 10 minutes twice, and washed with 2xSSC twice. Fluorescent fiducial beads (Molecular Probes, F-8809) were added at a 1:10,000 dilution in 2xSSC before imaging.

Secondary probes. Secondary (readout) probes (10 nM) were hybridized in secondary hybridization buffers (2xSSC, 20% formamide, and 10% dextran sulfate) to their primary targets for 30 minutes at 37 °C. Cells remained on the microscope stage during the hybridization. An objective heater was used to maintain the temperature at 37 °C. Cells were washed with secondary wash buffers (2xSSC, 20% formamide).

Fluidics and STORM Imaging. Multiple rounds of sequential labeling, washing, imaging, and bleaching were performed on an automated platform consisted of a fluidics setup and a STORM (stochastic optical reconstruction microscopy) microscope. The fluidics setup included a flow chamber (Bioptech FCS2), a peristaltic pump (Rainin Dynamax RP-1), and three computer-controlled 8-way valves (Hamilton MVP and Hamilton HVXM 8-5). This system allowed the automated integration of STORM movie collection and secondary hybridization.

The imaging buffer included, 50 mM Tris (pH 8) 10% (w/v) glucose, 1% βME (2-mercaptoethanol) or 25 mM MEA, with or without 2 mM 1,5-cyclooctadiene, and an oxygen scavenging system (0.5 mg/ml glucose oxidase (Sigma-Aldrich) and 40 microgram/ml catalase (Sigma-Aldrich)). A layer of mineral oil was used to seal the imaging buffer, preventing its acidification over the course of multiple hybridizations.

The STORM setup included an Olympus IX-71 inverted microscope configured for oblique incidence excitation. The samples were continuously illuminated with a 642-nm diode-pumped solid-state laser (VFL-P500-642; MPB communications). A 405-nm solid-state laser (Cube 405-100C; Coherent) was used for activation of dyes. Fluorescence was collected using an Olympus (UPlanSApo 100x, 1.4 NA) objective lens and passed through a custom dichroic, as well as a quad-view beam splitter. All movies were recorded using an EMCCD camera (Andor iAxon 897), imaging at 60Hz. The 512x256 field of view of the camera was split into separate 256x256 pixel movies prior to saving. The left half of this field of view contained the STORM data and the right half contained images of the fluorescent feducial beads. These latter movies were downsampled to 1 Hz prior to saving. During data acquisition, a home-built focus lock was used to maintain a constant focal plane. STORM movies included 20,000 to 30,000 frames in STORM buffer while the bleach movies included 10,000 frames in wash buffer.

Image Analysis – analysis of single-molecule localizations. Movies of single molecule localizations and fluorescent feducial beads were processed separately using a previously published single-emitter localization software.

Image Registration. The starting position of the beads from each round of hybridization were used to align movies from each round. The 2D autocorrelation between bead images of consecutive hybridizations followed by nearest-neighbor matching was used to match beads between images. The pair of beads with the most similar displacement vector were used to compute a rigid translation-rotation warp to align the beads. This alignment method is robust to samples in which multiple feducials are displaced or come detached and reattach during imaging.

Drift Correction. Drift during image acquisition was corrected using the trajectory of the feducial beads (recorded at 1 Hz). Bead positions were linked in each frame. The trajectory of the two beads that moved in the most correlated fashion was taken as the drift trajectory.

mRNA Cluster Calling. Localizations were first screened to be above a threshold number of photons (generally 2000) and required to be within 32 nm of 5 other localizations (parameters may be tuned). The remaining molecule localizations were binned in a 2D histogram of 10x10 nm bins (bin size is a variable parameter). All connected bins were taken to be part of a cluster (diagonal contacts are classified as connected). Clusters were required to have more than 80 total localizations across all hybridizations (variable parameter) to be called an mRNA cluster. The weighted centroids of these clusters from the 2D histogram were recorded as the mRNA positions.

A given cluster is recorded to be represented in an individual hybridization round if more than 9 localizations (variable parameter) are found within a 48 nm radius (variable parameter) of the centroid for that mRNA in each hybridization round.

Cluster Decoding. For each mRNA cluster, a codeword is readout, including "0"s for all the hybridization rounds in which less than the threshold number of localizations are found near the centroid and "1" s for the rounds where above threshold number of localizations are counted. The SECDED codebook decoded these as either perfect matches to target mRNA codewords, correctable errors which can be unambiguously mapped back to target mRNA, or uncorrectable errors, which differed by two or more letters from the words in the codebook.

Fig. 2A shows a STORM image of a cell. Fig. 2B shows a zoom in of the boxed region in Fig. 2A. Each dot indicates a localization. Localizations from different rounds of imaging are shown differently. Fig. 2C shows a representative cluster of localizations from the boxed region in Fig. 2B. The cluster shows localization signals from 4 different hybridizations. This cluster is a putative mRNA encoded with codeword [0 1 0 1 1 1 0 0].

Fig. 2D shows a reconstructed cell image of 14 genes after decoding and error correction. Each gene is shown differently. Fig. 2E shows measured gene expression for the 14 genes from the cell. Fig. 2F shows a comparison of transcript count with ensemble RNA-sequencing data. Fig. 2G shows correlation of transcript expression level between two cells detected using the described approach.

EXAMPLE 4

The following examples are generally directed to multiplexed single-molecule imaging with error-robust encoding allowing for simultaneous measurements of thousands of RNA species in single cells. In general, knowledge of the expression profile and spatial landscape of RNAs in individual cells is essential for understanding the rich repertoire of cellular behaviors. The following examples reports various techniques directed to single-molecule imaging approaches that allow the copy numbers and spatial localizations of thousands of RNA species to be determined in single cells. Some of these techniques are called Multiplexed Error-Robust Fluorescence in Situ Hybridization or "MERFISH."

Using error-robust encoding schemes to combat single-molecule labeling and detection errors, these examples demonstrated the imaging of hundreds to thousands of unique RNA species in hundreds of individual cells. Correlation analysis of the ~10^4 to ~10^6 pairs of genes allowed constraints on gene regulatory networks, prediction of novel functions for many unannotated genes, and identification of distinct spatial distribution patterns of RNAs that correlate with properties of the encoded proteins.

System-wide analyses of the abundance and spatial organization of RNAs in single cells promise to transform understanding in many areas of cell and developmental biology, such as the mechanisms of gene regulation, the heterogeneous behavior of cells, and the development and maintenance of cell fate. Single-molecule fluorescence *in situ* hybridization (smFISH) has emerged as a powerful tool for studying the copy number and spatial organization of RNAs in single cells either in isolation or in their native tissue context. Taking advantage of its ability to map the spatial distributions of specific RNAs with high resolution, smFISH has revealed the importance of subcellular RNA localization in diverse processes such as cell migration, development, and polarization. In parallel, the ability of smFISH to precisely measure the copy numbers of specific RNAs without amplification bias has allowed quantitative measurement of the natural fluctuations in gene expression, which has in turn elucidated the regulatory mechanisms that shape such fluctuations and their role in a variety of biological processes.

However, application of the smFISH approach to many systems-level questions remains limited by the number of RNA species that can be simultaneously measured in single cells. State-of-the-art efforts using combinatorial labeling by either color-based barcodes or sequential hybridization have enabled simultaneous measurements of 10-30 different RNA species in individual cells, yet many interesting biological questions would benefit from the measurement of hundreds to thousands of RNAs within a single cell, which are not achievable using such techniques. For example, analysis of how the expression profile of such a large number of RNAs vary from cell to cell and how these variations correlate among different genes could be used to systematically identify co-regulated genes and map regulatory networks; knowledge of the subcellular organizations of numerous RNAs and their correlations could help elucidate molecular mechanisms underlying the establishment and maintenance of many local cellular structures; and RNA profiling of individual cells in native tissues could allow *in situ* identification of cell type.

The following examples generally discuss certain techniques called MERFISH, which are highly multiplexed smFISH imaging methods that substantially increase the number of RNA species that can be simultaneously imaged in single cells by using combinatorial labeling and sequential imaging with error-robust encoding schemes. These examples demonstrate this multiplexed imaging approach by simultaneously measuring 140 RNA species using an encoding scheme that can both detect and correct errors and 1001 RNA species using an encoding scheme that can detect but not correct errors. It should be understood that these numbers are by way of exemplification only, not limitation. Correlation analyses of the copy number variations and spatial distributions of these genes allowed us to identify groups of genes that are co-regulated and groups of genes that share similar spatial distribution patterns inside the cell.

Combinatorial labeling with error-robust encoding schemes. Combinatorial labeling that identifies each RNA species by multiple (N) distinct signals offers a route to rapidly increase the number of RNA species that can be probed simultaneously in individual cells (Fig. 5A). However, this approach to scaling up the throughput of smFISH to the systems scale faces a significant challenge because not only does the number of addressable RNA species increases exponentially with N, but the detection error rates also grow exponentially with N (Figs. 5B-5D). Imagine a conceptually simple scheme to implement combinatorial labeling, where each RNA species is encoded with a N-bit binary word and the sample is probed with N corresponding rounds of hybridization, each round targeting only the subset of RNAs that should read '1' in the corresponding bit (Fig. 11). N rounds of hybridization

would allow 2^N-1 RNA species to be probed. With just 16 hybridizations, over 64,000 RNA species, which should cover the entire human transcriptome including both messenger RNAs (mRNAs) and non-coding RNAs, could be identified (Fig. 5B; upper symbols). However, as *N* increases, the fraction of RNAs properly detected (the calling rate) would rapidly decrease and, more troublingly, the fraction of RNAs that are identified as incorrect species (the misidentification rate) would rapidly increase (Fig. 5C, lower symbols; Fig. 5D, upper symbols). With realistic error rates per hybridization (measured below), the majority of RNA molecules would be misidentified after 16 rounds of hybridizations!

To address this challenge, error-robust encoding schemes were designed, in which only a subset of the 2^N-1 words separated by a certain Hamming distance were used to encode RNAs. In a codebook where the minimum Hamming distance is 4 (HD4 code), at least four bits must be read incorrectly to change one code word into another (Fig. 12A). As a result, every single-bit error produces a word that is uniquely close to a single code word, allowing such errors to be detected and corrected (Fig. 12B). Double-bit errors produce words with an equal Hamming distance of 2 from multiple code words and, thus, can be detected but not corrected (Fig. 12C). Such a code should substantially increase the calling rate and reduce the misidentification rate (Fig. 5C and 5D, middle symbols). To further account for the fact that it is more likely to miss a hybridization event (an 1-->0 error) than to misidentify a background spot as an RNA (an 0-->1 error) in smFISH measurements, a modified HD4 (MHD4) code was designed, in which the number of '1' bits were kept both constant and relatively low, only four per word, to reduce error and avoid biased detection. This MHD4 code should further increase the calling rate and reduce the misidentification rate (Fig. 5C, upper symbols; Fig. 5D, lower symbols).

In addition to the error considerations, several practical challenges have also made it difficult to probe a large number of RNA species, such as the high cost of the massive number of fluorescently labeled FISH probes needed and the long time required to complete many rounds of hybridization. To overcome these challenges, in this example, a two-step labeling scheme was designed to encode and readout cellular RNAs (Fig. 5E). First, cellular RNAs were labeled with a set of encoding probes (also called primary probes), each probe comprising a RNA targeting sequence and two flanking readout sequences. Four of the *N* unique readout sequences were assigned to each RNA species based on the MHD4 code word of the RNA. Second, these *N* readout sequences were identified with complementary FISH probes, the readout probes (also called secondary probes) via *N* rounds of hybridization and imaging, each round using a unique readout probe. To increase the signal to background

ratio, every cellular RNA was labeled with ~192 encoding probes. Because each encoding probe contained two of the four readout sequences associated with that RNA (Fig. 5E), a maximum of ~96 readout probes can bind to each cellular RNA per hybridization round. To generate the massive number of encoding probes required, they were amplified from array-derived oligonucleotide pools containing tens of thousands of custom sequences using an enzymatic amplification process comprising *in vitro* transcription followed by reverse transcription (Fig. 13, see below regarding probe synthesis). This two-step labeling approach significantly diminished the total hybridization time for an experiment: it was found that efficient hybridization to the readout sequences took only 15 minutes whereas efficient direct hybridization to cellular RNA required more than 10 hours.

Fig. 5 describes MERFISH, a highly multiplexed smFISH approach using combinatorial labeling and error-robust encoding. Fig. 5A shows a schematic depiction of the identification of multiple RNA species in N rounds of imaging. Each RNA species is encoded with a N-bit binary word and during each round of imaging, only the subset of RNAs that should read '1' in the corresponding bit emit signal. Figs. 5B-5D show the number of addressable RNA species (Fig. 5B), the rate at which these RNAs are properly identified (calling rate) (Fig. 5C), and the rate at which RNAs are incorrectly identified as a different RNA species (misidentification rate) (Fig. 5D) plotted as a function of the number of bits (N) in the binary words encoding RNA. In Figs. 5B and 5D, the upper dots are a simple binary code that includes all 2^N-1 possible binary words; the middle dots are the HD4 code where the Hamming distance separating words is 4; and the lower dots are the modified HD4 (MHD4) code where the number of '1' bits are kept at four. These are reversed in Fig. 5C.

The calling and misidentification rates are calculated with per bit error rates of 10% for the 1-->0 error and 4% for the 0-->1 error. Fig. 5E is as schematic diagram of the implementation of a MHD4 code for RNA identification. Each RNA species is first labeled with ~192 encoding probes that convert the RNA into a unique combination of readout sequences (Encoding hyb). These encoding probes each contain a central RNA targeting region flanked by two readout sequences, drawn from a pool of N different sequences, each associated with a specific hybridization round. Encoding probes for a specific RNA species contain a unique combination of four of the N readout sequences, which correspond to the four hybridization rounds where this RNA should read '1'. N subsequent rounds of hybridization with the fluorescent readout probes were used to probe the readout sequences (hyb 1, hyb 2, ..., hyb N). The bound probes were inactivated by photobleaching between

successive rounds of hybridization. For clarity, only one possible pairing of the readout sequences is depicted here for the encoding probes; however, all possible pairs of the four readout sequences are used at the same frequency and distributed randomly along each cellular RNA in the actual experiments.

Fig. 11 shows a schematic description of a combinatorial labeling approach based on a simple binary code. In a conceptually simple labeling approach, 2^N-*1* different RNA species can be uniquely encoded with all *N*-bit binary words (excluding the word with all '0's). In each hybridization round, FISH probes that are targeted to all RNA species that have a '1' in the corresponding bit are included. To increase the ability to discriminate RNA spots from background, each RNA is addressed with multiple FISH probes per hybridization round. Signal from the bound probes is extinguished before the next round of hybridization. This process continues for all *N* hybridization rounds (hyb 1, hyb 2, ...), and all 2^N-*1* RNA species can be identified by the unique on-off pattern of fluorescence signals in each hybridization round.

Fig. 12 shows schematic descriptions of Hamming distance and its use in the identification and correction of errors. Fig. 12A is a schematic representation of a Hamming distance of 4. Figs. 12B and 12C are schematics showing the ability of an encoding scheme with Hamming distance 4 to correct single-bit errors (Fig. 12B) or detect but not correct double-bit errors (Fig. 12C). Arrows highlight bits at which the indicated words differ. Two code words are separated by a Hamming distance of 4 if one of the words has to flip four bits from '1' to '0' or '0' to '1' to convert into the other word. Single-bit error correction is possible because if a measured word differs from a legitimate code word by only one bit, it is most likely an error that arises from misreading this code word, since the code words of all the other RNA species will differ from the measured word by at least three bits. In this case, the measured word can be corrected to a code word that differs by only one bit. If a measured word differs from a legitimate code word by two bits, this measured word can still be identified as an error, but correction is no longer possible since more than one legitimate code word differs from this measured word by two bits.

Fig. 13 shows production of the library of encoding probes. An array-synthesized complex oligopool, containing ~100k sequences, is used as a template for the enzymatic amplification of the encoding probes for different experiments. Each template sequence in the oligopool contains a central target region that can bind to a cellular RNA, two flanking readout sequences, and two flanking index primers. In the first step, the required template molecules for a specific experiment are selected and amplified with an indexed PCR reaction.

To allow amplification via *in vitro* transcription, a T7 promotor is added to the PCR products during this step. In the second step, RNA is amplified from these template molecules via *in vitro* transcription. In the third step, this RNA is reverse transcribed back into DNA. In the final step, the template RNA is removed via alkaline hydrolysis, leaving only the desired ssDNA probes. This protocol produces ~2 nmol of complex pools of encoding probes containing ~20,000 different sequences for the 140-gene experiments or ~100,000 different sequences for the 1001-gene experiments.

EXAMPLE 5

This example illustrates the measurement of 140 genes with MERFISH using a 16-bit MHD4 Code. To test the feasibility of this error-robust, multiplexed imaging approach, this example uses a 140-gene measurement on human fibroblast cells (IMR90) using a 16-bit MHD4 code to encode 130 RNA species while leaving 10 code words as misidentification controls (Figs. 20A to 20H). After each round of hybridization with the fluorescent readout probes, cells were imaged by conventional wide-field imaging with an oblique-incidence illumination geometry. Fluorescent spots corresponding to individual RNAs were clearly detected and were then efficiently extinguished via a brief photobleaching step (Fig. 6A). The sample was stable throughout the 16 rounds of iterative labeling and imaging. The change in the number of fluorescent spots from round to round matched the predicted change based on the relative abundances of RNA species targeted in each round derived from bulk sequencing, and a systematic decreasing trend with increasing number of hybridization rounds was not observed (Fig. 14A). The average brightness of the spots varied from round to round with a standard deviation of 40%, likely due to different binding efficiencies of the readout probes to the different readout sequences on the encoding probes (Fig. 14B). Only a small, systematic decreasing trend in the spot brightness with increasing hybridization rounds was observed, which was on average 4% per round (Fig. 14B).

Next, binary words were constructed from the observed fluorescent spots based on their on-off patterns across the 16 hybridization rounds (Fig. 6B-6D). If the word exactly matched one of the 140 MHD4 code words (exact matches) or differed by only one bit (error-correctable matches), it was assigned to the corresponding RNA species (Fig. 6D). Within the single cell depicted in Figs. 6A and 6B, more than 1500 RNA molecules corresponding to 87% of the 130 encoded RNA species were detected after error correction (Fig. 6E). Similar observations were made in ~400 cells from 7 independent experiments. On average, ~4 times as many RNA molecules and ~2 times as many RNA species were detected per cell after error correction as compared with the values obtained before error correction (Fig. 15).

Two types of errors can occur in the copy number measurement of each RNA species: 1) Some molecules of this RNA species are not detected, leading to a drop in calling rate, and 2) some molecules from other RNA species are misidentified as this RNA species. To assess the extent of misidentification, the 10 misidentification control words were utilized, i.e., code words that were not associated with any cellular RNA. Although matches to these control words were observed, they occurred far less frequently than the real RNA-encoding words: 95% of the 130 RNA-encoding words were counted more frequently than the median count for these control words. Moreover, it was typically found that the ratio of the number of exact matches to the number of matches with one-bit errors for a real RNA-encoding word was substantially higher than the same ratios observed for the misidentification controls, as expected (Figs. 16A and 16B). Using this ratio as a measure of the confidence in RNA identification, it was found that 91% of the 130 RNA species had a confidence ratio greater than the maximum confidence ratio observed for the misidentification controls (Fig. 6F), demonstrating a high accuracy of RNA identification. Subsequent analyses were conducted only on these 91% of genes.

To estimate the calling rate, the error-correction ability of the MHD4 code was utilized to determine the 1-->0 error rates (10% on average) and 0-->1 error rates (4% on average) for each hybridization round (Figs. 16C and 16D). Using these error rates, an ~80% calling rate for individual RNA species after error correction was estimated, i.e. ~80% of the fluorescent spots corresponding to a RNA species were decoded correctly (Fig. 16E). It is noted that although the remaining 20% of spots contributed to a loss in detection efficiency, most of them did not cause species misidentification because they were decoded as double-bit error words and discarded.

To test for potential technical bias in these measurements, the same 130 RNAs species were probed with a different MHD4 codebook by shuffling the code words among different RNA species (Figs. 20A to 20H) and changing the encoding probe sequences. Measurements with this alternative code gave similar misidentification and calling rates (Fig. 17). The copy numbers of individual RNA species per cell measured with these two codebooks showed excellent agreement with a Pearson correlation coefficient of 0.94 (Fig. 6G), indicating that the choice of encoding scheme did not bias the measured counts.

In order to validate the copy numbers derived from the MERFISH experiments, conventional smFISH measurements were performed on 15 of the 130 genes, selected from the full measured abundance range of three orders of magnitude. For each of these genes, both the average copy number and the copy number distribution across many cells agreed

quantitatively between the MERFISH and conventional smFISH measurements (Fig. 18A and 18B). The ratio of the copy numbers determined by these two approaches was 0.82 +/- 0.06 (mean +/- SEM across the 15 measured RNA species, Fig. 18B), which agreed with the estimated 80% calling rate for the multiplexed imaging approach. The quantitative match between this ratio and the estimated calling rate over the full measured abundance range additionally supports the assessment that the misidentification error was low. Given that the agreement between the MERFISH and conventional smFISH results extended to the genes at the lowest measured abundance (<1 copy per cell, Fig. 18B), it was estimated that the measurement sensitivity was at least 1 copy per cell.

As a final validation, the abundance of each RNA species averaged over hundreds of cells was compared to those obtained from a bulk RNA sequencing measurement that were performed on the same cell line. The imaging results correlated remarkably well with bulk sequencing results with a Pearson correlation coefficient of 0.89 (Fig. 6H).

Fig. 6 shows simultaneous measurement of 140 RNA species in single cells using MERFISH with a 16-bit MHD4 code. Fig. 6A shows images of RNA molecules in an IMR90 cell after each hybridization round (hyb 1 – hyb 16). The image after photobleaching (bleach 1) demonstrated efficient removal of fluorescent signals between hybridizations. Fig. 6B shows the localizations of all detected single molecules in this cell colored based on their measured binary words. Inset: the composite fluorescent image of the 16 hybridization rounds for the boxed sub-region with numbered circles indicating potential RNA molecules. A circle indicates an unidentifiable molecule, the binary word of which does not match any of the 16-bit MHD4 code words even after error correction. Fig. 6C shows fluorescent images from each round of hybridization for the boxed sub-region in Fig. 6B with circles indicating potential RNA molecules. Fig. 6D shows corresponding words for the spots identified in Fig. 6C. Crosses represent the corrected bits. Fig. 6E shows the RNA copy number for each gene observed without (lower) or with (higher) error correction in this cell. Fig. 6F shows the confidence ratio measured for the 130 RNA species (left) and the 10 misidentification control words (right) normalized to the maximum value observed from the misidentification controls (dashed line). Fig. 6G is a scatter plot of the average copy number of each RNA species per cell measured with two shuffled codebooks of the MHD4 code. The Pearson correlation coefficient is 0.94 with a p-value of 1×10^{-53}. The dashed line corresponds to the $y = x$ line. Fig. 6H is a scatter plot of the average copy number of each RNA species per cell versus the abundance determined by bulk sequencing in fragments per kilobase per million reads

(FPKM). The Pearson correlation coefficient between the logarithmic abundances of the two measurements was 0.89 with a p-value of $3x10^{-39}$.

Fig. 14 shows the number and average brightness of the fluorescent spots detected in the 16 rounds of hybridization before and after photobleaching. Fig. 14A shows the number of fluorescent spots observed per cell before (higher) and after (lower) photobleaching as a function of hybridization round averaged across all measurements with the first 16-bit MHD4 code. Photobleaching reduces the number of fluorescent spots by two or more orders of magnitude. Hybridization rounds without lower bars represent rounds in which no molecules were observed after bleaching. Also depicted is the expected change in the number of fluorescent spots from round to round (circles) predicted based on the relative abundances of the RNA species targeted in each hybridization round derived from bulk RNA sequencing. The average discrepancy between the observed and predicted number of spots for each hybridization is only 15% of the mean number of spots. This discrepancy does not systematically increase with the number of hybridization rounds. Fig. 14B shows the average brightness of the identified fluorescent spots in each hybridization round averaged across all measurements with the first 16-bit MHD4 code both before (upper) and after (lower) photobleaching. Brightness varies by 40% (standard deviation) across different hybridization rounds. The variation pattern is reproducible between experiments with the same code, likely due to differences in the binding efficiency of the readout probes to the different readout sequences. There is a small systematic trend of decrease in the brightness with increasing hybridization rounds, which is on average 4% per round. Photobleaching extinguishes fluorescence to a level similar to that of the autofluorescence of the cell.

Fig. 15 shows error correction substantially increases the numbers of RNA molecules and RNA species detected in individual cells. Fig. 15A shows a histogram of the ratio of the total number of molecules detected per cell with error correction to the number measured without error correction. Fig. 15B is a histogram of the total number of RNA species detected in each cell with error correction to that without error correction. Both ratios are determined for ~200 cells and the histograms are constructed from these ratios.

Fig. 16 shows characterization of the misidentification and calling rates of RNA species for the 140-gene experiments using a specific 16-bit MHD4 code. Fig. 16A shows the number of measured words exactly matching the code word corresponding to FLNA, represented by the bar in the center of the circle, and the number of measured words with one-bit error compared to the code word of FLNA, represented by the 16 bars on the circle. Fig. 16B is the same as Fig. 16A, but for a code word that was not assigned to any RNA, i.e.,

a misidentification control word. The solid lines connect the exact match to one-bit error words that are generated by 1-->0 errors. Based on the observation that the ratio of the number of exact matches to the number of error-correctable matches for a real RNA-encoding word was typically substantially higher than the same ratios observed for the misidentification controls, this ratio was defined as a confidence ratio for RNA identification. The confidence ratio measured for all 130 RNA species (center bar) and 10 misidentification control words not assigned to any RNA (outer bars) using this 16-bit MHD4 code is show in Fig. 6F. Fig. 16C and 16D show the average error rates for the 1-->0 error (Fig. 16C) and 0-->1 error (Fig. 16D) for each hybridization round. Fig. 16E shows the calling rate for each RNA species estimated from the 1-->0 and 0-->1 error rates. Genes are sorted from left to right based on the measured abundance, which spans three orders of magnitude. The calling rates are largely independent of the abundance of the gene.

Fig. 17 shows characterization of the misidentification and calling rates for a second 16-bit MHD4 code. In this second encoding scheme, the 140 code words were shuffled among different RNA species and changed the encoding probe sequences. Fig. 17A shows the normalized confidence ratio measured for the 130 RNA species (left) and the 10 misidentification control words not assigned to any RNA (right). The normalized confidence ratio is determined the same way as in Fig. 6F. Figs. 17B and 17C show the average error rates determined for the 1-->0 error (Fig. 17B) and 0-->1 error (Fig. 17C) for each hybridization round. Fig. 17D shows the calling rate determined for each RNA species estimated from the 1-->0 and 0-->1 error rates. Genes are sorted from left to right based on the measured abundance.

Fig. 18 shows a comparison of the MERFISH measurements with conventional smFISH results for a subset of genes. Fig. 18A shows the distributions of RNA copy numbers in single cells for three example genes KIAA1199, DYNC1H1, and LMTK2 in the high, medium, and low abundance ranges, respectively. Lighter bars: distributions constructed from ~400 cells in the 140-gene measurements using the MHD4 codes. Darker bars: distributions constructed from ~100 cells in the conventional smFISH measurements. Fig. 18B shows a comparison of the average RNA copy numbers per cell measured in the 140-gene experiments using the MHD4 codes to those determined by conventional smFISH for 15 genes. The average ratio of the copy number measured using the MHD4 measurements to that measured using conventional smFISH was 0.82 +/- 0.06 (mean +/- SEM across 15 genes). The dashed line corresponds to the $y = x$ line.

Figs. 20A to 20H show two different codebooks for the 140-gene experiments. The specific code words of the 16-bit MHD4 code assigned to each RNA species studied in the two shuffles of the 140-gene experiment. The "Genes" columns contain the name of the gene. The "Codewords" columns contain the specific binary word assigned to each gene.

EXAMPLE 6

This example is generally directed to high-throughput analysis of cell-to-cell variation in gene expression. The MERFISH approach allows parallelization of measurements of many individual RNA species and co-variation analysis between different RNA species. In this example, the parallelization aspect was first illustrated by examining the cell-to-cell variation in the expression level of each of the measured genes (Fig. 7A). To quantify the measured variation, the Fano factors, defined as the ratio of the variance to the mean RNA copy number, were computed for all measured RNA species. The Fano factors substantially deviated from 1, the value expected for a simple Poisson process, for many genes and exhibited an increasing trend with the mean RNA abundance (Fig. 7B). This trend of increasing Fano factors with mean RNA abundance can be explained by changes in the transcription rate and/or promoter off-switching rates but not by changes in the promoter on-switching rate.

Moreover, several RNA species were identified with substantially larger Fano factors than this average trend. For example, it was found that SLC5A3, CENPF, MKI67, TNC and KIAA1199 displayed Fano factor values substantially higher than those of the other genes expressed at similar abundance levels. The high variability of some of these genes can be explained by their association with the cell cycle. For example, two of these particularly 'noisy' genes MKI67 and CENPF were both annotated as cell-cycle related genes, and based on their bimodal expression (Fig. 7C), it is proposed that their transcription is strongly regulated by the cell cycle. Other high-variability genes did not show the same bimodal expression patterns and are not known to be associated with the cell cycle.

Analysis of co-variations in the expression levels of different genes can reveal which genes are co-regulated and elucidate gene regulatory pathways. At the population level, such analysis often requires the application of external stimuli to drive gene expression variation; hence, correlated expression changes can be observed among genes that share common regulatory elements influenced by the stimuli. At the single-cell level, one can take advantage of the natural stochastic fluctuations in gene expression for such analysis and can thus study multiple regulatory networks without having to stimulate each of them

individually. Such co-variation analysis can constrain regulatory networks, suggest new regulatory pathways, and predict function for unannotated genes based on associations with co-varying genes.

This approach was applied to the 140-gene measurements and the ~10,000 pairwise correlation coefficients describing how the expression levels of each pair of genes co-varied from cell to cell were examined. Many of the highly variable genes showed tightly correlated or anti-correlated variations (Fig. 7C). To better understand the correlations for all gene pairs, a hierarchical clustering approach was adopted to organize these genes based on their correlation coefficients (Fig. 7D). From the cluster tree structure, seven groups of genes with substantially correlated expression patterns were identified (Fig. 7D). Within each of the seven groups, every gene showed significantly stronger average correlation with other members of the group than with genes outside the group. To further validate and understand these groups, gene ontology (GO) terms enriched in each of these seven groups were identified. Notably, the enriched GO terms within each group shared similar functions and were largely unique to each group (Fig. 7E), validating the notion that the observed co-variation in expression reflects some commonalities in the regulation of these genes.

This example describes two of these groups as illustrative examples. The predominant GO terms associated with Group 1 were terms associated with the extracellular matrix (ECM) (Figs. 7D and 7E). Notable members of this group included ECM components, such as FBN1, FBN2, COL5A, COL7A and TNC, and glycoproteins linking the ECM and cell membranes, such as VCAN and THBS1. The group also included an unannotated gene, KIAA1199, which may predict to play a role in ECM metabolism based on its association with this cluster. Indeed, this gene has recently been identified as an enzyme involved in the regulation of hyaluronan, a major sugar component of the ECM.

Group 6 contained many genes that encode vesicle transport proteins and proteins associated with cell motility (Figs. 7D and 7E). The vesicle transport genes included microtubule motors and related genes DYNC1H, CKAP1, and factors associated with vesicle formation and trafficking, like DNAJC13 and RAB3B. Again, an unannotated gene, KIAA1462, was found within this cluster. Based on its strong correlation with DYNC1H1 and DNAJC13, it is predicted that this gene may be involved in vesicle transport. The cell motility genes in this group included actin-binding proteins like AFAP1, SPTAN1, SPTBN1, and MYH10, and genes involved in the formation of adhesion complexes, like FLNA and FLNC. Several GTPase-associated factors involved in the regulation of cell motility, attachment and contraction also fell into this group, including DOCK7, ROCK2, IQGAP1,

PRKCA, and AMOTL1. The observation that some cell motility genes correlated with vesicle transport genes is consistent with the role of vesicle transport in cell migration. An additional interesting feature of group 6 is that a subset of these genes, in particular those related to cell motility, were anti-correlated with members of the ECM group discussed above (Fig. 7D). This anti-correlation may reflect regulatory interactions that mediate switching of cells between adherent and migratory states.

Fig. 7 shows cell-to-cell variations and pairwise correlations for the RNA species determined from the 140 gene measurements. Fig. 7A shows a comparison of gene expression levels in two individual cells. Fig. 7B shows Fano factors for individual genes. Error bars represent standard error of the mean determined from 7 independent data sets. Fig. 7C shows Z-scores of the expression variations of four example pairs of genes showing correlated (top two) or anti-correlated (bottom two) variation for 100 randomly selected cells. The Z-score is defined as the difference from the mean normalized by the standard deviation. Fig. 7D is a matrix of the pairwise correlation coefficients of the cell-to-cell variation in expression for the measured genes, shown together with the hierarchical clustering tree. The seven groups identified by a specific threshold on the cluster tree (dashed line) are indicated by the black boxes in the matrix and lines on the tree, with grey lines on the tree indicating ungrouped genes. Different threshold choices on the cluster tree could be made to select either smaller subgroups with tighter correlations or larger super-groups containing more weakly coupled subgroups. Two of the seven groups are enlarged on the right. Fig. 7E shows enrichment of 30 selected, statistically significantly enriched GO terms in the seven groups. Enrichment refers to the ratio of the fraction of genes within a group that have the specific GO term to the fraction of all measured genes having that term. Not all of the GO terms presented here are in the top 10 list.

EXAMPLE 7

This example illustrates mapping spatial distributions of RNAs. As an imaging based approach, MERFISH also allowed the investigation of the spatial distributions of many RNA species simultaneously. Several patterns emerged from the visual inspection of individual genes, with some RNA transcripts enriched in the perinuclear region, some enriched in the cell periphery, and some scattered throughout the cell (Fig. 8A). To identify genes with similar spatial distributions, the correlation coefficients for the spatial density profiles of all pairs of RNA species were determined, and these RNAs were organized based on the pairwise correlations again using the hierarchical clustering approach. The correlation coefficient matrix showed groups of genes with correlated spatial organizations, and the two

most notable groups with the strongest correlations are indicated in Fig. 8B. Group I RNAs appeared enriched in the perinuclear region whereas group II RNAs appeared enriched near the cell peripheral region (Fig. 8C). Quantitative analysis of the distances between each RNA molecule and the cell nucleus or the cell periphery indeed confirmed this visual impression (Fig. 8D).

Group I contained genes encoding extracellular proteins such as FBN1, FBN2 and THSB1, secreted proteins such as PAPPA, and integral membrane proteins such as LRP1 and GPR107. These proteins have no obvious commonalities in function. Rather a GO analysis showed significant enrichment for location terms, such as extracellular region, basement membrane, or perivitelline space (Fig. 8E). To reach these locations, proteins must pass through the secretion pathway, which often requires translation of mRNA at the endoplasmic reticulum (ER). Thus, it is proposed that the spatial pattern that were observed for these mRNAs reflects their co-translational enrichment at the ER. The enrichment of these mRNAs in the perinuclear region (Figs. 8C and 8D, lighter shading), where the rough ER resides, supports this conclusion.

Group II contained genes encoding the actin-binding proteins, including filamins FLNA and FLNC, talin TLN1, and spectrins SPTAN1 and SPTBN1; the microtubule-binding protein CKAP5; and the motor proteins MYH10 and DYNC1H1. This group was enriched with GO terms such as cortical actin cytoskeleton, actin filament binding, and cell-cell adherens junction (Fig. 8E). Beta-actin mRNA may be enriched near the cell periphery in fibroblasts as are mRNAs that encode members of the actin-binding Arp2/3 complex. The enrichment of group II mRNAs in the peripheral region of the cells (Figs. 8C and 8D) suggests that the spatial distribution of the Group II genes might be related to the distribution of actin cytoskeleton mRNAs.

Fig. 8 shows distinct spatial distributions of RNAs observed in the 140-gene measurements. Fig. 8A shows examples of the spatial distributions observed for four different RNA species in a cell. Fig. 8B is a matrix of the pairwise correlation coefficients describing the degree with which the spatial distributions of each gene pair is correlated, shown together with the hierarchical clustering tree. Two strongly correlating groups are indicated by the black boxes on the matrix and shading on the tree. Fig. 8C shows the spatial distributions of all RNAs in the two groups in two example cells. Lighter symbols: group I genes; darker symbols: group II genes. Fig. 8D shows average distances for genes in group I and genes in group II to the cell edge or the nucleus normalized to the average distances for

all genes. Error bars represent SEM across 7 data sets. Fig. 8E shows enrichment of GO terms in each of the two groups.

EXAMPLE 8

This example illustrates measuring 1001 genes with a 14-bit MHD2 code. This example further increases the throughput of MERFISH measurements by simultaneously imaging ~1000 RNA species. This increase could be achieved with the MHD4 code by increasing the number of bits per code word to 32 while maintaining the number of '1' bits per word at four (Fig. 5B). While the stability of the samples across many hybridization rounds (Fig. 14) suggests that such an extension is potentially feasible, an alternative approach is shown here that did not require an increase in the number of hybridizations by relaxing the error correction requirement but keeping the error detection capability. For example, by reducing the Hamming distance from 4 to 2, all 14-bit words could be used that contain four '1' bits to encode 1001 genes and these RNAs were probed with only 14 rounds of hybridization. However, because a single error can produce a word equally close to two different code words, error correction is no longer possible for this modified Hamming-distance-2 (MHD2) code. Hence, it was expected that the calling rate would be lower and the misidentification rate to be higher with this encoding scheme.

To evaluate the performance of this 14-bit MHD2 code, 16 of the 1001 possible code words were set aside as misidentification controls and used the remaining 985 words to encode cellular RNAs. Among these 985 RNAs included 107 RNA species probed in the 140-gene experiments as an additional control. The 1001-gene experiments were performed in IMR90 cells using a similar procedure as described above. To allow all encoding probes to be synthesized from a single 100,000-member oligopool, the number of encoding probes per RNA species was reduced to ~94. Fluorescent spots corresponding to individual RNA molecules were again detected in each round of hybridization with the readout probes and, based on their on-off patterns, these spots were decoded into RNA (Figs. 9A, 19A and 19B). 430 RNA species were detected in the cell shown in Fig. 9A, and similar results were obtained in ~200 imaged cells in 3 independent experiments.

As expected, the misidentification rate of this scheme was higher than that of the MHD4 code. 77% of all real RNA words were detected more frequently than the median count for the misidentification controls instead of the 95% value observed in the MHD4 measurements. Using the same confidence ratio analysis as described above, it was found that 73% (instead of 91% for the MHD4 measurements) of the 985 RNA species were measured with a confidence ratio larger than the maximum value observed for the

misidentification controls (Fig. 19C). RNA copy numbers measured from these 73% RNA species showed excellent correlation with the bulk RNA sequencing results (Pearson correlation coefficient r = 0.76; Fig. 9B, black). It is worth noting that the remaining 27% of the genes still exhibit good, albeit lower, correlation with the bulk RNA sequencing data (r = 0.65; Fig. 9B, red), but the conservative measure of excluding them from further analysis was taken.

The lack of an error correction capability also decreased the calling rate of each RNA species: When comparing the 107 RNA species common in both the 1001-gene and 140-gene measurements, it was found that the copy numbers per cell of these RNA species were lower in the 1001-gene measurements (Figs. 9C and 19D). The total count of these RNAs per cell was ~1/3 of that observed in the 140-gene measurements. Thus the lack of error correction in the MHD2 code produced a ~3-fold decrease in the calling rate, which is consistent with the ~4-fold decrease in calling rate observed for the MHD4 code when error correction was not applied. As expected from the quantitative agreement between 140-gene measurements and conventional smFISH results, comparison of the 1001-gene measurements with conventional smFISH results for 10 RNA species also indicated a ~3-fold drop in calling rate (Fig. 18C). Despite the expected reduction in calling rate, the good correlations found between the copy numbers observed in the 1001-gene measurements and those observed in the 140-gene measurements, as well as in conventional smFISH and bulk RNA sequencing measurements, indicates that the relative abundance of these RNAs can be quantified with the MHD2 encoding scheme.

Simultaneously imaging ~1000 genes in individual cells substantially expanded the ability to detect co-regulated genes. Fig. 10A shows the matrix of pairwise correlation coefficients determined from the cell-to-cell variations in the expression levels of these genes. Using the same hierarchical clustering analysis as described above, ~100 groups of genes with correlated expression were identified. Remarkably, nearly all of these ~100 groups showed statistically significant enrichment of functionally related GO terms (Fig. 10B). These included some of the groups identified in the 140-gene measurements, such as the group associated with cell replication genes and the group associated with cell motility genes (Figs. 10A and 10B, groups 7 and 102), as well as many new groups. The groups identified here included 46 RNA species lacking any previous GO annotations, for which function based on their group association may be hypothesized. For example, KIAA1462 is part of the cell motility group, as also shown in the 140-gene experiments, suggesting a potential role of this gene in cell motility (Fig. 10A, group 102). Likewise, KIAA0355 is part

of a new group enriched in genes associated with heart development (Fig. 10A, group 79), and C17orf70 is part of a group associated with ribosomal RNA processing (Fig. 10A, group 22). Using these groupings, cellular functions for 61 transcription factors and other partially annotated proteins of unknown functions may be hypothesized. For example, the transcription factors Z3CH13 and CHD8 are both members of the cell motility group, suggesting their potential role in the transcriptional regulation of cell motility genes.

Fig. 9 shows simultaneous measurements of 1001 genes in single cells using MERFISH with a 14-bit MHD2 code. Fig. 9A shows the localizations of all detected single molecules in a cell colored based on their measured binary words. Inset: the composite, false-colored fluorescent image of the 14 hybridization rounds for the boxed sub-region with numbered circles indicating potential RNA molecules. Circles indicate unidentifiable molecules, the binary words of which do not match any of the 14-bit MHD2 code words. Images of individual hybridization round are shown in Fig. 19A. Fig. 9B is a scatter plot of the average copy number per cell measured in the 1001-gene experiments versus the abundance measured via bulk sequencing. The upper symbols are for the 73% of genes detected with confidence ratios higher than the maximum ratio observed for the misidentification controls. The Pearson correlation coefficient is 0.76 with a p-value of 3×10^{-133}. The lower symbols are for the remaining 27% of genes. The Pearson correlation coefficient is 0.65 with a p-value of 3×10^{-33}. Fig. 9C is a scatter plot of the average copy number for the 107 genes shared in both the 1001-gene measurement with the MHD2 code and the 140-gene measurement with the MHD4 code. The Pearson correlation coefficient is 0.89 with a p-value of 9×10^{-30}. The dashed line corresponds to the $y = x$ line.

Fig. 10 shows co-variation analysis of the RNA species measured in the 1001-gene measurements. Fig. 10A is a matrix of all pairwise correlation coefficients of the cell-to-cell variation in expression for the measured genes shown with the hierarchical clustering tree. The ~100 identified groups of correlated genes are indicated by shading on the tree. Zoom in of four of the groups described in the text are shown on the right. Fig. 10B is an enrichment of 20 selected, statistically significantly enriched GO terms in the four groups.

Fig. 19 shows decoding and error assessment of the 1001-gene experiments. Fig. 19A shows images of the boxed sub-region of the cell in Fig. 9A for each of 14 hybridization rounds. The final panel is a composite image of these 14 rounds. Circles indicate fluorescent spots that have been identified as potential RNA molecules. Some circles in the composite image indicate unidentifiable molecules, the binary words of which do not match any of the 14-bit MHD2 code words. Fig. 9B shows the corresponding binary word for each of the

spots identified in Fig. 9A with the RNA species to which it is decoded. 'unidentified' implies that the measured binary word does not match any of the 1001 code words. Fig. 19C shows the normalized confidence ratios measured for the 985 RNA species (left) and the 16 misidentification control words not targeted to any RNA (right). The normalized confidence ratio is defined as in Fig. 6F. Fig. 19D shows a histogram of the reduction in detected abundance of 107 genes present in both the 1001-gene experiments and the 140-gene experiments. "Fold decrease in copy number" is defined as the average number of RNA molecules per cell for each species measured in the 140-gene experiments divided by the corresponding average number measured in the 1001-gene experiments.

Fig. 18C is a comparison of the average RNA copy numbers per cell measured in the 1001-gene experiments using the MHD2 code to those determined by conventional smFISH for 10 genes. The average ratio of the copy number measured using the MHD2 measurements to that measured using conventional smFISH was 0.30 +/- 0.05 (mean +/- SEM across 10 genes). The dashed line corresponds to the $y = x$ line and the dotted line corresponds to the $y = 0.30x$ line.

EXAMPLE 9

The above examples illustrate a highly multiplexed detection scheme for systems-level RNA imaging in single cells. Using combinatorial labeling, sequential hybridization and imaging, and two different error-robust encoding schemes, either 140 or 1001 genes in hundreds of individual human fibroblast cells were simultaneously imaged. Of the two encoding schemes presented here, the MHD4 code is capable of both error detection and error correction, and hence can provide a higher calling rate and a lower misidentification rate than the MHD2 code, which instead can only detect but cannot correct errors. MHD2, on the other hand, provides a faster scaling of the degree of multiplexing with the number of bits than MHD4. Other error-robust encoding schemes can also be used for such multiplexed imaging, and experimenters can set the balance between detection accuracy and ease of multiplexing based on the specific requirements of the experiments.

By increasing the number of bits in the code words, it should be possible to further increase the number of detectable RNA species using MERFISH with, for example, a MHD4 or MHD2 code. For example, using the MHD4 code with 32 total bits and four or six '1' bits would increase the number of addressable RNA species to 1,240 or 27,776, respectively. The latter is the approximate scale of the human transcriptome. The predicted misidentification and calling rates are still reasonable for the 32-bit MHD4 code (shown in Figs. 5C and 5D for the MHD4 code with four '1' bits and similar rates were calculated for the MHD4 code with

six '1' bits). If more accurate measurements are desired, an additional increase in the number of bits would allow the use of encoding schemes with a Hamming distance greater than 4, further enhancing the error detection and correction capability. While an increase in the number of bits by adding more hybridization rounds would increase the data collection time and potentially lead to sample degradation, these problems could be mitigated by utilizing multiple colors to readout multiple bits in each round of hybridization.

As the degree of multiplexing is increased, it is important to consider the potential increase in the density of RNAs that need to be resolved in each round of imaging. Based on the imaging and sequencing results, it can be estimated that including the whole transcriptome of the IMR90 cells would lead to a total RNA density of ~200 molecules/micrometer^3. Using the current imaging and analysis methods, 2-3 molecules/ micrometer^3 per hybridization round could be resolved, which would reach a total RNA density of ~20 molecules/micrometer^3 after 32 rounds of hybridization. This density should allow all but the top 10% most expressed genes to be imaged simultaneously or a subset of genes with even higher expression levels to be included. By utilizing more advanced image analysis algorithms to better resolve overlapping images of individual molecules, such as compressed sensing, it is possible to extend the resolvable density by ~4-fold and thus allow all but the top 2% most expressed genes to be imaged all together.

These examples have illustrated the utility of the data derived from highly multiplexed RNA imaging by using co-variation and correlation analysis to reveal distinct sub-cellular distribution patterns of RNAs, to constrain gene regulatory networks, and to predict functions for many previously unannotated or partially annotated genes with unknown functions. Given its ability to quantify RNAs across a wide range of abundances without amplification bias while preserving native context, systems and methods such as MERFISH will allow many applications of *in situ* transcriptomic analyses of individual cells in culture or complex tissues.

EXAMPLE 10

Following are various materials and methods used in the above examples.

Probe design. Each RNA species in the target set was randomly assigned a binary code word either from all 140 possible code words of the 16-bit MHD4 code or from all 1001 possible code words of the 14-bit MHD2 code.

Array-synthesized oligopools were used as templates to make the encoding probes. The template molecule for each encoding probe contained three components: i) a central targeting sequence for *in situ* hybridization to the target RNA, ii) two flanking readout

sequences designed to hybridize each of two distinct readout probes, and iii) two flanking primer sequences to allow enzymatic amplification of the probes (Fig. 13). The readout sequences were taken from the 16 possible readout sequences, each corresponding to one hybridization round. The readout sequences were assigned to the encoding probes such that for any RNA species each of the 4 readout sequences were distributed uniformly along the length of the target RNA and appeared at the same frequency. Template molecules for the 140-gene library also included a common 20-nucleotide (nt) priming region between the first PCR primer and the first readout sequence. This priming sequence was used for the reverse transcription step described below.

Multiple experiments were embedded in a single array-synthesized oligopool, and PCR was used to selectively amplify only the oligos required for a specific experiment. Primer sequences for this indexed PCR reaction were generated from a set of orthogonal 25-nt sequences. These sequences were trimmed to 20 nt and selected for i) a narrow melting temperature range (70 °C to 80 °C), ii) the absence of consecutive repeats of 3 or more identical nucleotides, and iii) the presence of a GC clamp, i.e. one of the two 3' terminal bases must be G or C. To further improve specificity, these sequences were then screened against the human transcriptome using BLAST+, and primers with 14 or more contiguous bases of homology were eliminated. Finally, BLAST+ was again used to identify and exclude primers that had an 11-nt homology region at the 3' end of any other primer or a 5-nt homology region at the 3' end of the T7 promoter. The forward primer sequences (Primer 1) were determined as described above, whereas the reverse primers each contain a 20-nt sequence as described above plus a 20-nt T7 promoter sequence to facilitate amplification via *in vitro* transcription (Primer 2). The primer sequences used in the 140-gene and 1001-gene experiments are listed below.

Table 2

Experiment Name	Primer 1 Sequence (Index Primer 1)	Primer 2 Sequence (T7 promoter plus the reverse complement of Index Primer 2)
140-gene Codebook 1	GTTGGTCGGCACTTGGGTGC (SEQ ID NO: 18)	TAATACGACTCACTATAGGGAAAGCCGGTTCATCCGGTGG (SEQ ID NO: 21)
140-gene Codebook 2	CGATGCGCCAATTCCGGTTC (SEQ ID NO: 19)	TAATACGACTCACTATAGGGTGATCATCGCTCGCGGGTTG (SEQ ID NO: 22)
1001-gene	CGCGGGCTATATGCGAACC	TAATACGACTCACTATAGGGCGTGGAGG

G (SEQ ID NO: 20) GCATACAACGC (SEQ ID NO: 23)

30-nt-long readout sequences were created by concatenating fragments of the same orthogonal primer set generated above by combining one 20-nt primer with a 10-nt fragment of another. These readout sequences were then screened, using BLAST+, for orthogonality with the index primer sequences and other readout sequences (no more than 11 nt of homology) and for potential off-target binding sites in the human genome (no more than 14 nt of homology). Fluorescently labeled readout probes with sequences complementary to the readout sequences were used to probe these readout sequences, one in each hybridization round. All used readout probes sequences are listed below.

Table 3

Bit	Readout probes	
1	CGCAACGCTTGGGACGGTTCCAATCGGATC/3Cy5Sp/	SEQ ID NO: 24
2	CGAATGCTCTGGCCTCGAACGAACGATAGC/3Cy5Sp/	SEQ ID NO: 25
3	ACAAATCCGACCAGATCGGACGATCATGGG/3Cy5Sp/	SEQ ID NO: 26
4	CAAGTATGCAGCGCGATTGACCGTCTCGTT/3Cy5Sp/	SEQ ID NO: 27
5	GCGGGAAGCACGTGGATTAGGGCATCGACC/3Cy5Sp/	SEQ ID NO: 28
6	AAGTCGTACGCCGATGCGCAGCAATTCACT/3Cy5Sp/	SEQ ID NO: 29
7	CGAAACATCGGCCACGGTCCCGTTGAACTT/3Cy5Sp/	SEQ ID NO: 30
8	ACGAATCCACCGTCCAGCGCGTCAAACAGA/3Cy5Sp/	SEQ ID NO: 31
9	CGCGAAATCCCCGTAACGAGCGTCCCTTGC/3Cy5Sp/	SEQ ID NO: 32
10	GCATGAGTTGCCTGGCGTTGCGACGACTAA/3Cy5Sp/	SEQ ID NO: 33
11	CCGTCGTCTCCGGTCCACCGTTGCGCTTAC/3Cy5Sp/	SEQ ID NO: 34
12	GGCCAATGGCCCAGGTCCGTCACGCAATTT/3Cy5Sp/	SEQ ID NO: 35
13	TTGATCGAATCGGAGCGTAGCGGAATCTGC/3Cy5Sp/	SEQ ID NO: 36
14	CGCGCGGATCCGCTTGTCGGGAACGGATAC/3Cy5Sp/	SEQ ID NO: 37
15	GCCTCGATTACGACGGATGTAATTCGGCCG/3Cy5Sp/	SEQ ID NO: 38
16	GCCCGTATTCCCGCTTGCGAGTAGGGCAAT/3Cy5Sp/	SEQ ID NO: 39

The readout probes used for the 140-gene libraries were probes 1 through 16. The readout probes used for the 1001-gene experiment were probes 1 through 14. /3Cy5Sp/ indicates a 3' Cy5 modification.

To design the central targeting sequences of the encoding probes, the abundance of different transcripts in IMR90 cells using Cufflinks v2.1, total RNA data from the ENCODE project, and human genome annotations from Gencode v18 were complied. Probes were designed from gene models corresponding to the most abundant isoform using OligoArray2.1 with the following constraints: the target sequence region is 30-nt long; the melting temperatures of the hybridized region of the probe and cellular RNA target is greater than 70 °C; there is no cross hybridization targets with melting temperatures greater than 72 °C; there

is no predicted internal secondary structures with melting temperatures greater than 76 °C; and there is no contiguous repeats of 6 or more identical nucleotides. Melting temperatures were adjusted to optimize the specificity of these probes and minimize secondary structure while still producing sufficient numbers of probes for the libraries. To decrease computational cost, isoforms were divided into 1-kb regions for probe design. Using BLAST+, all potential probes that mapped to more than one cellular RNA species were rejected. Probes with multiple targets on the same RNA were kept.

For each gene in the 140-gene experiments, 198 putative encoding probe sequences were generated by concatenating the appropriate index primers, readout sequences, and targeting regions as shown in Fig. 13. To address the possibility that concatenation of these sequences introduced new regions of homology to off-target RNAs, BLAST+ was used to screen these putative sequences against all human rRNA and tRNA sequences as well as highly expressed genes (genes with FPKM > 10,000). Probes with greater than 14 nt of homology to rRNAs or tRNAs or greater than 17 nt of homology to highly expressed genes were removed. After these cuts, there were ~192 (with a standard deviation of 2) probes per gene for both MHD4 codebooks used in the 140-gene experiments. The same protocol for the 1001-gene experiments was used, as follows: Starting with 96 putative targeting sequences per gene, ~94 (with a standard deviation of 6) encoding probes per gene were obtained after these additional homology cuts. The number of encoding probes per RNA was decreased for the 1001-gene experiments so that these probes could be synthesized from a single 100,000-member oligopool as opposed to two separate pools. Each encoding probe was designed to contain two of the four readout sequences associated with each code word, hence only half of the bound encoding probes can bind readout probe during any given hybridization round. ~192 or ~94 encoding probes per RNA were used to obtain high signal-to-background ratios for individual RNA molecules. The number of encoding probes per RNA could be substantially reduced but still allow single RNA molecules to be identified. In addition, increasing the number of readout sequences per encoding probe or using optical sectioning methods to reduce the fluorescence background may allow further reduction in the number of the encoding probes per RNA.

Two types of misidentification controls were designed. The first control (blank words) were not represented with encoding probes. The second type of control (no-target words) had encoding probes that were not targeting any cellular RNA. The targeting regions of these probes were composed of random nucleotide sequences subject to the same constraints used to design the RNA targeting sequences described above. Moreover, these

random sequences were screened against the human transcriptome to ensure that they contain no significant homology (>14-nt) to any human RNA. The 140-gene measurements contained 5 blank words and 5 no-target words. The 1001-gene measurements contained 11 blank words and 5 no-target words.

Probe synthesis. The encoding probes were synthesized using the following steps, and this synthesis protocol is illustrated in Fig. 13.

Step 1: The template oligopool (CustomArray) was amplified via limited-cycle PCR on a Bio-Rad CFX96 using primer sequences specific to the desired probe set. To facilitate subsequent amplification via *in vitro* transcription, the reverse primer contained the T7 promoter. All primers were synthesized by IDT. This reaction was column purified (Zymo DNA Clean and Concentrator; D4003).

Step 2: The purified PCR products were then further amplified ~200-fold and converted into RNA via a high yield *in vitro* transcription according to the manufacturer's instructions (New England Biolabs, E2040S). Each 20 microliter reaction contained ~1 microgram of template DNA from above, 10 mM of each NTP, 1x reaction buffer, 1x RNase inhibitor (Promega RNasin, N2611) and 2 microliters of the T7 polymerase. This reaction was incubated at 37 °C for 4 hours to maximize yield. This reaction was not purified before the following steps.

Step 3: The RNA products from the above *in vitro* transcription reaction were then converted back into DNA via a reverse transcription reaction. Each 50 microliter reaction contained the unpurified RNA produce from Step 2 supplemented with 1.6 mM of each dNTP, 2 nmol of a reverse transcription primer, 300 units of Maxima H- reverse transcriptase (Thermo Scientific, EP0751), 60 units of RNasin, and a final 1x concentration of the Maxima RT buffer. This reaction was incubated at 50 °C for 45 minutes, and the reverse transcriptase was inactivated at 85 °C for 5 minutes. The templates for the 140-gene libraries contain a common priming region for this reverse transcription step; thus, a single primer was used for this step when creating these probes. Its sequence was CGGGTTTAGCGCCGGAAATG (SEQ ID NO: 40). A common priming region was not included for the 1001-gene library; thus, the reverse transcription was conducted with the forward primer: CGCGGGCTATATGCGAACCG (SEQ ID NO: 20).

Step 4: To remove the template RNA, 20 microliters of 0.25 M EDTA and 0.5 N NaOH was added to the above reaction to selectively hydrolyze RNA, and the sample was incubated at 95 °C for 10 minutes. This reaction was then immediately purified by column purification using a 100-microgram-capacity column (Zymo Research, D4030) and the Zymo

Oligo Clean and Concentrator protocol. The final probes were eluted in 100 microliters of RNase-free deionized water, evaporated in a vacuum concentrator, and then resuspended in 10 μL of encoding hybridization buffer (see below). Probes were stored at -20 °C. Denaturing poly-acrylamide gel electrophoresis and absorption spectroscopy were used to confirm the quality of the probes and revealed that this probe synthesis protocol converts 90-100% of the reverse-transcription primer into full length probe and of the probe that is constructed, 70-80% is recovered during the purification step.

Fluorescently labeled readout probes have sequences complementary to the readout sequences described above and a Cy5 dye attached at the 3' end. These probes were obtained from IDT and HPLC purified.

Sample preparation and labeling with encoding probes. Human primary fibroblasts (American Type Culture Collection, IMR90) were used in this work. These cells are relatively large and flat, facilitating wide-field imaging without the need for optical sectioning. Cells were cultured with Eagle's Minimum Essential Medium. Cells were plated on 22-mm, #1.5 coverslips (Bioptechs, 0420-0323-2) at 350,000 cells/coverslip and incubated at 37 °C with 5% CO_2 for 48-96 hours within petri dishes. Cells were fixed for 20 minutes in 4% paraformaldehyde (Electron Microscopy Sciences, 15714) in 1x phosphate buffered saline (PBS; Ambion, AM9625) at room temperature, reduced for 5 minutes with 0.1% w/v sodium borohydride (Sigma, 480886) in water to reduce background fluorescence, washed three times with ice-cold 1x PBS, permeabilized for 2 minutes with 0.5% v/v Triton (Sigma, T8787) in 1x PBS at room temperature, and washed three times with ice cold 1x PBS.

Cells were incubated for 5 minutes in encoding wash buffer comprising 2x saline-sodium citrate buffer (SSC) (Ambion, AM9763), 30% v/v formamide (Ambion, AM9342), and 2 mM vanadyl ribonucleoside complex (NEB, S1402S). 10 microliters of 100 micromolar(140-gene experiments) or 200 micromolar (1001-gene experiments) encoding probes in encoding hybridization buffer was added to the cell-containing coverslip and spread uniformly by placing another coverslip on top of the sample. Samples were then incubated in a humid chamber inside a 37 °C-hybridization oven for 18-36 hours. Encoding hybridization buffer is composed of encoding wash buffer supplemented with 1 mg/mL yeast tRNA (Life Technologies, 15401-011) and 10% w/v dextran sulfate (Sigma, D8906-50G).

Cells were then washed with primary encoding wash buffer, incubated at 47 °C for 10 minutes, and this wash was repeated for a total of three times. A 1:1000 dilution of 0.2-micrometer-diameter carboxylate-modified orange fluorescent beads (Life Technologies, F-8809) in 2xSSC was sonicated for 3 minutes and then incubated with the sample for 5

minutes. The beads were used as fiducial markers to align images obtained from multiple successive rounds of hybridization, as described below. The sample was washed once with 2xSSC, and then post-fixed with 4% v/v paraformaldehyde in 2xSSC at room temperature for 30 minutes. The sample was then washed three times with 2xSSC and either imaged immediately or stored for no longer than 12 hours at 4 °C prior to imaging. All solutions were prepared as RNase-free.

MERFISH imaging with multiple successive rounds of hybridization. The sample coverslip was assembled into a Bioptech's FCS2 flow chamber, and the flow through this chamber was controlled via a home-built fluidics system composed of three computer-controlled 8-way valves (Hamilton, MVP and HVXM 8-5) and a computer-controlled peristaltic pump (Rainin, Dynamax RP-1). The sample was imaged on a home-built microscope constructed around an Olympus IX-71 body and a 1.45 NA, 100x oil immersion objective and configured for oblique incidence excitation. The objective was heated to 37 °C with a Bioptechs objective heater. Constant focus was maintained throughout the imaging process with a home-built, auto-focusing system. Illumination was provided at 641 nm, 561 nm, and 405 nm using solid state lasers (MPB communications, VFL-P500-642; Coherent, 561-200CWCDRH; and Coherent, 1069413/AT) for excitation of the Cy5-labeled readout probes, the fiducial beads, and nuclear counterstains, respectively. These lines were combined with a custom dichroic (Chroma, zy405/488/561/647/752RP-UF1) and the emission was filtered with a custom dichroic (Chroma, ZET405/488/561/647-656/752m). Fluorescence was separated with a QuadView (Photometrics) using the dichroics T560lpxr, T650lpxr, 750dcxxr (Chroma) and the emission filters ET525/50m, WT59550m-2f, ET700/75m, HQ770lp (Chroma) and imaged with an EMCCD camera (Andor, iXon-897). The camera was configured so that a pixel corresponds to 167 nm in the sample plane. The entire system was fully automated, so that imaging and fluid handling were performed for the entire experiment without user intervention.

Sequential hybridization, imaging, and bleaching proceeded as follows. 1 mL of 10 nM of the appropriate fluorescently labeled readout probe in readout hybridization buffer (2xSSC; 10% v/v formamide; 10% w/v dextran sulfate, and 2 mM vanadyl ribonucleoside complex) was flown across the sample, flow was stopped, and the sample was incubated for 15 minutes. Then 2 mL of readout wash buffer (2xSSC, 20% v/v formamide; and 2 mM vanadyl ribonucleoside complex) was flown across the sample, flow was stopped, and the sample was incubated for 3 minutes. 2 mL of imaging buffer comprising 2xSSC, 50 mM TrisHCl pH 8, 10% w/v glucose, 2 mM Trolox (Sigma-Aldrich, 238813), 0.5 mg/mL glucose

oxidase (Sigma-Aldrich, G2133), and 40 microgram/mL catalase (Sigma-Aldrich, C30) was flown across the sample. Flow was then stopped, and then approximately 75 to 100 regions were exposed to ~25 mW 642-nm and 1 mW of 561-nm light and imaged. Each region was 40 micrometers by 40 micrometers. The laser powers were measured at the microscope backport. Because the imaging buffer is sensitive to oxygen, the ~50 mL of imaging buffer used for a single experiment was made fresh at the beginning of the experiment and then stored under a layer of mineral oil throughout the measurement. Buffer stored in this fashion was stable for more than 24 hours.

After imaging, the fluorescence of the readout probes was extinguished via photobleaching. The sample was washed with 2 mL of photobleaching buffer (2xSSC and 2 mM vanadyl ribonucleoside complex), and each imaged region of the sample was exposed to 200 mW of 641-nm light for 3 s. To confirm the efficacy of this photobleaching treatment, imaging buffer was reintroduced, and the sample was imaged as described above.

The above hybridization, imaging, and photobleaching process was repeated either 16 times for the 140-gene measurements using the MHD4 code or 14 times for the 1001-gene measurements using the MHD2 code. An entire experiment was typically completed in ~20 hours.

Following completion of imaging, 2 mL of a 1:1000 dilution of Hoescht (ENZ-52401) in 2xSSC was flown through the chamber to label the nuclei of the cells. The sample was then washed immediately with 2 mL of 2xSSC followed by 2 mL of imaging buffer. Each region of the sample was then imaged once again with ~1 mW of 405-nm light.

Because cells were imaged using wide-field imaging with oblique-incidence illumination, without optical sectioning and z-scanning, the fraction of individual RNA species that was outside the axial range of the imaging geometry was quantified for 6 different RNA species using conventional smFISH. For this purpose, these cells were optically sectioned by collecting stacks of images at different focal depths through the entire depth of the cells. The images were aligned in consecutive focal planes and then computed for each cell the fraction of RNAs that were detected in the three-dimensional stack but not in the basal focal plane. It was found that only a small fraction, 15% +/- 1% (Mean +/- SEM across six different RNA species) of RNA molecules were outside the imaging range of a fixed focal plane without z-scanning. These measurements also confirmed that the excitation geometry illuminated the full depth of the cells. Any optical sectioning technique could be employed in MERFISH to allow the imaging of RNAs in thicker cells or tissues.

Construction of measured words. Fluorescent spots were identified and localized in each image using a multi-Gaussian-fitting algorithm assuming a Gaussian with a uniform width of 167 nm. This algorithm was used to allow partially overlapping spots to be distinguished and individually fit. RNA spots were distinguished from background signal, i.e. signal arising from probes bound non-specifically, by setting the intensity threshold required to fit a spot with this software. Due to variation in the brightness of spots between rounds of hybridization, this threshold was adjusted appropriately for each hybridization round to minimize the combined average of the 1-->0 and 0-->1 error rates across all hybridization rounds (140-gene measurements) or to maximize the ratio of the number of measured words with four '1' bits to those with three or five '1' bits (1001-gene measurements). The location of the fiducial beads was identified in each frame using a faster single-Gaussian fitting algorithm.

Images of the same sample region in different rounds of hybridization were registered by rotating and translating the image to align the two fiducial beads within the same image that were most similar in location after a coarse initial alignment via image correlation. All images were aligned to a coordinate system established by the images collected in the first round of hybridization. The quality of this alignment was determined from the residual distance between five additional fiducial beads, and alignment error was typically ~20 nm.

Fluorescence spots in different hybridization rounds were connected into a single string, corresponding to a potential RNA molecule, if the distance between spots was smaller than 1 pixel (167 nm). For each string of spots, the on-off sequence of fluorescent signals in all hybridization rounds were used to assign a binary word to the potential RNA molecule, in which '1' was assigned to the hybridization rounds that contained a fluorescent signal above threshold and '0' was assigned to the other hybridization rounds. Measured words were then decoded into RNA species using the 16-bit MHD4 code or the 14-bit MHD2 code discussed above. In the case of the 16-bit MHD4 code, if the measured binary word matched the code word of a specific RNA perfectly or differed from the code word by one single bit, it was assigned to that RNA. In the case of the 14-bit MHD2 code, only if the measured binary word matched the code word of a specific RNA perfectly, was it assigned to that RNA. To determine the copy number per cell, the number of each RNA species was counted in individual cells within each 40 micrometer by 40 micrometer imaging area. It is noted that this number accounts for the majority but not all RNA molecules within a cell because a fraction of the cell could be outside the imaging area or focal depth. Tiling images of

adjacent areas and adjacent focal planes could be employed to improve the counting accuracy.

In the 140-gene experiments, some regions of the cell nucleus occasionally contained too much fluorescence signal to properly identify individual RNA spots. In the 1001-gene experiments, the cell nucleus generally contained too much fluorescent signal to allow identification of individual RNA molecules. These bright regions were excluded from all subsequent analysis. This work focuses on mRNAs, which are enriched in the cytoplasm. To estimate the fraction of mRNAs missed by excluding the nucleus region, conventional smFISH was used to quantify the fraction of molecules found inside the nucleus for six different mRNAs species. It was found that only 5% +/- 2% (Mean +/- SEM across six RNA species) of these RNA molecules are found in the nucleus. Employment of super-resolution imaging and/or optical sectioning could potentially allow individual molecules in these dense nucleus regions to be identified, which would be particularly useful for probing those non-coding RNAs that are enriched in the nucleus.

smFISH measurements of individual genes. Pools of 48 fluorescently-labeled (Quasar 670) oligonucleotide probes per RNA were purchased from Biosearch Technologies. 30-nt probe sequences were taken directly from a random subset of the targeting regions used for the multiplexed measurements. Cells were fixed and permeabilized as described above. 10 microliters of 250 nM oligonucleotide probes in encoding hybridization buffer (described above) was added to the cell-containing coverslip and spread uniformly by placing another coverslip on top of the sample. Samples were then incubated in a humid chamber inside a 37 ^{o}C-hybridization oven for 18 hours. Cells were then washed with encoding wash buffer (described above) at 37 ^{o}C for 10 minutes, and this wash was repeated for a total of three times. The sample was then washed three times with 2xSSC and imaged in imaging buffer using the same imaging geometry as described above for MERFISH.

Bulk RNA sequencing. Total RNA was extracted from IMR90 cells cultured as above using the Zymo Quick RNA MiniPrep kit (R1054) according to the manufacturer's instructions. polyA RNA was then selected (NEB; E7490), and a sequencing library was constructed using the NEBNext Ultra RNA library preparation kit (NEB; E7530), amplified with custom oligonucleotides, and 150-bp reads were obtained from on a MiSeq. These sequences were aligned to the human genome (Gencode v18) and isoform abundance was computed with cufflinks.

Calculation of the predicted scaling and error properties of different encoding schemes. Analytic expressions were derived for the dependence of the number of possible

code words, the calling rate, and the misidentification rate on *N*. The calling rate is defined as the fraction of RNA molecules that are properly identified. The misidentification rate is defined as the fraction of RNA molecules that are misidentified as a wrong RNA species. For encoding schemes with an error-detection capability, the calling rate and misidentification rate does not add up to 1 because a fraction of the molecules not called properly can be detected as errors and discarded and, hence, not misidentified as a wrong species. These calculations assume that the probability of misreading bits is constant for all hybridization rounds but differs for the 1-->0 and 0-->1 errors. Experimentally measured average 1-->0 and 0-->1 error rates (10% and 4% respectively) were used for the estimates shown in Figs. 5B-5D. For simplicity, the word corresponding to all '0's was not removed from calculations.

For the simple binary encoding scheme in which all possible *N*-bit binary words are assigned to unique RNA species, the number of possible code words is 2^N. The number of words that could be used to encode RNA is actually $2^N - 1$ because the code word '00...0' does not contain detectable fluorescence in any hybridization round, but for simplicity the word corresponding to all '0's was not removed from subsequent calculations. The error introduced by this approximation is negligible. For any given word with *m* '1's and *N*–*m* '0's the probability of measuring that word without error (the fraction of RNAs that is properly called) is:

$$(1-p_1)^m(1-p_0)^{N-m}, \tag{1}$$

where p_1 is 1-->0 error rate and p_0 is 0-->1 error rate per bit. Because different words in this simple binary encoding scheme can have different numbers of '1' bits, the calling rate for different words will differ if $p_1 \neq p_0$. The average calling rate, reported in Fig. 5C, was determined from the weighted average of the value of Eq. (1) for all words. This weighted average is:

$$\frac{1}{2^N}\sum_{m=0}^{N}\binom{N}{m}(1-p_1)^m(1-p_0)^{N-m}, \tag{2}$$

where $\binom{N}{m}$ is the binomial coefficient and corresponds to the number of words with *m* '1' bits in this encoding scheme. Since in this encoding scheme every error produces a binary word that encodes a different RNA, the average misidentification rate for this encoding scheme, reported in Fig. 5D, follows directly from (2):

$$1-\frac{1}{2^N}\sum_{m=0}^{N}\binom{N}{m}(1-p_1)^m(1-p_0)^{N-m}. \quad (3)$$

To calculate the scaling and error properties of the extended Hamming distance 4 (HD4) code, the generator matrix for the desired number of data bits using standard methods was first created. The generator matrix determines the specific words that are present in a given encoding scheme and was used to directly determine the number of encoded words as a function of the number of bits. In this encoding scheme, the calling rate corresponds to the fraction of words measured without error as well as the fraction of words measured with a single-bit error. For code words with *m* '1' bits, this fraction is determined by the following expression:

$$(1-p_1)^m(1-p_0)^{N-m}+m\,p_1{}^1(1-p_1)^{m-1}(1-p_0)^{N-m}+(N-m)\,p_0{}^1(1-p_1)^m(1-p_0)^{N-m-1} \quad (4)$$

where the first term is the probability of not making any errors, the second term corresponds to the total probability of making one 1-->0 error at any of the *m* '1' bits without making any other 0-->1 errors, and the final term corresponds to the total probability of making one 0-->1 error at any of the *N-m* '0' bits without making any 1-->0 errors. Because the number of '1' bits can differ between words in this encoding scheme, the average calling rate reported in Fig. 5C was computed from a weighted average over Eq. (4) for different values of *m*. The weight for each term was determined from the number of words that contain *m* '1' bits as determined from the generator matrix described above.

Because RNA-encoding words are separated by a minimum Hamming distance of 4, at least 4 errors are required to switch one word into another. If error correction is applied, then 3 or 5 errors could also convert one RNA into another. Thus, the misidentification rate from all possible combinations of 3-bit, 4-bit and 5-bit errors was estimated for code words with *m* '1' bits. Technically, >5-bit errors could also convert one RNA into another, but the probability of making such errors is negligible because of the small per-bit error rate. This expression was approximated with:

$$\begin{aligned}
&\sum_{i=0}^{4}\binom{m}{i}\binom{N-m}{4-i}p_1^i p_0^{4-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(4-i)}+\\
&\sum_{i=0}^{3}\binom{m}{i}\binom{N-m}{3-i}p_1^i p_0^{3-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(3-i)}+\\
&\sum_{i=0}^{5}\binom{m}{i}\binom{N-m}{5-i}p_1^i p_0^{5-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(5-i)}+
\end{aligned} \quad (5)$$

The first sum corresponds to all of the ways in which exactly four mistakes can be made. Similarly, the second and third sums correspond to all of the ways in which exactly three or

five mistakes can be made. Eq. (5) provides an upper bound for the misidentification rate because not all three, four, or five bit errors produce a word that matches or would be corrected to another legitimate word. Again because the number of '1' bits can differ between words, the average misidentification rate reported in Fig. 5D is calculated as a weighted average of Eq. (5) over the number of words that have *m* '1' bits.

To generate the MHD4 code where the number of '1' bits for each code word is set to 4, the HD4 codes were first generated as described above, and then all code words that did not contain four '1's were removed. The calling rate of this code, reported in Fig. 5C, was directly calculated from Eq. (4) but with m = 4 because all code words in this code have four '1' bits. The misidentification rate of this code, reported in Fig. 5D, was calculated by modifying Eq. (5) with the following considerations: (i) the number of '1' bits, m, was set to 4 and (ii) errors that produce words that do not contain three, four, or five '1' bits were excluded. Thus, the expression in Eq. (5) was simplified to

$$\begin{aligned}
&\binom{4}{2}\binom{N-4}{2}p_1^2 p_0^2(1-p_1)^2(1-p_0)^{N-6} + \\
&\binom{4}{1}\binom{N-4}{2}p_1^1 p_0^2(1-p_1)^3(1-p_0)^{N-6} + \binom{4}{2}\binom{N-4}{1}p_1^2 p_0^1(1-p_1)^2(1-p_0)^{N-5} \\
&\binom{4}{2}\binom{N-4}{3}p_1^2 p_0^3(1-p_1)^2(1-p_0)^{N-7} + \binom{4}{3}\binom{N-4}{2}p_1^3 p_0^2(1-p_1)^1(1-p_0)^{N-6}
\end{aligned} \tag{6}$$

Again, this expression is an upper bound on the actual misidentification rate because not all words with four '1's are valid code words.

Estimates of the 1-->0 and 0-->1 error rates for each hybridization round. To compute the probability of misreading a bit at a given hybridization round, the error correcting properties of the MHD4 code were used. Briefly, the probabilities of 1-->0 or 0-->1 errors were derived in the following way. Let the probability of making an error at the *i*th bit, i.e. *i*th hybridization round, be p_i and the actual number of RNA molecules of the given species be *A*, then the number of exact matches for this RNA will be $W_E = A\prod_{i=1}^{16}(1-p_i)$ and the number of one-bit error corrected matches for this RNA corresponding to errors at the *i*th bit will be $W_i = A\frac{p_i}{(1-p_i)}\prod_{j=1}^{16}(1-p_j)$. The p_i can be directly derived from the ratio: $W_i/W_E = \frac{p_i}{(1-p_i)}$. This ratio assumes that the one-bit error-corrected counts were only generated from single-bit errors from the correct word and that multi-error contamination from other RNA words is negligible. Given that the error rate per hybridization round is small and that it takes at least

three errors to convert one RNA-encoding word into a word that would be misidentified as another RNA, the above approximation should be a good one.

To compute the average 1-->0 or 0-->1 error probabilities for each of the 16 hybridization rounds, the above approach was used to calculate the per-bit error rates for each bit of every gene, and these errors were sorted based on whether they correspond to a 1-->0 or a 0-->1 error, and the average of these errors for each bit weighted by the number of counts observed for the corresponding gene was taken.

Estimates of the calling rate for individual RNA species from actual imaging data. With the estimates of the 1-->0 or 0-->1 error probabilities for each round of hybridization as determined above, it is possible to estimate the calling rate for each RNA based on the specific word used to encoded it. Specifically, the fraction of an RNA species that is called correctly is determined by

$$\prod_{i=1}^{N}(1-p_i)+\sum_{j=1}^{N}\frac{p_j}{(1-p_j)}\prod_{i=1}^{N}(1-p_i), \qquad (7)$$

where the first term represent the probability of observing an exact match of the code word and the second term represent the probability of observing an error-corrected match (i.e. with one-bit error). The values of the per-bit error rate p_i for each RNA species are determined by the specific code word for that RNA and the measured 1-->0 or 0-->1 error rates for each round of hybridization. If the code word of the RNA contains a '1' in the *i*th bit, then p_i is determined from the 1-->0 error rate for the *i*th hybridization round; if the word contains a '0' in the *i*th bit, p_i is determined from the 0-->1 error rate for the *i*th hybridization round.

Hierarchical clustering analysis of the co-variation in RNA abundance. Hierarchical clustering of the co-variation in gene expression for both the 140-gene and 1001-gene experiments was conducted as follows. First, the distance between every pair of genes was determined as 1 minus the Pearson correlation coefficient of the cell-to-cell variation of the measured copy numbers of these two RNA species, both normalized by the total RNA counted in the cell. Thus, highly correlated genes are 'closer' to one another and highly anti-correlated genes are 'further' apart. An agglomerative hierarchical cluster tree was then constructed from these distances using the Unweighted Pair Group Method with Arithmetic mean (UPGMA). Specifically, starting with individual genes, hierarchical clusters were constructed by identifying the two clusters (or individual genes) that are closest to one another according to the arithmetic mean of the distances between all inter-cluster gene pairs. The pairs of clusters (or individual genes) with the smallest distance are then grouped

together and the process is repeated. The matrix of pairwise correlations was then sorted based on the order of the genes within these trees.

Groups of genes with substantial co-variations were identified by selecting a threshold on the hierarchical cluster tree (indicated by the dashed lines in Figs. 7D and 10A) that produced approximately 10 groups of genes each of which contains at least 4 members for the 140-gene experiments or approximately 100 groups each of which contains at least 3 members for the 1001-gene experiments. It is noted that one can change the threshold in order to identify either more tightly coupled smaller groups or larger groups with relatively loose coupling.

A probability value for the confidence that a gene belongs to a specific group was determined by computing the difference between the average correlation coefficient between that gene and all other members of that group and the average correlation coefficient between that gene and all other measured genes outside that group. The significance (p-value) of this difference was determined with the student's t-test.

Because hierarchical clustering is inherently a one-dimensional analysis, i.e. any given genes can only be a member of a single group, this analysis does not allow all correlated gene groups to be identified. Higher dimension analysis, such as principal component analysis or k-means clustering, could be used to identify more co-varying gene clusters.

Analysis of RNA spatial distributions. To identify genes that have similar spatial distributions, each of the measured cells was subdivided into 2x2 regions and calculated the fraction of each RNA species present in each of these bins. To control for the fact that some regions of the cell naturally contain more RNA than others, the enrichment was calculated for each gene , i.e., the ratio of the observed fraction in a given region for a given RNA species to the average fraction observed for all genes in that same region. For each pair of RNA species, the Pearson correlation coefficient of the region-to-region variation in enrichment of these two RNA species for each cell was determined and the correlation coefficients were averaged over ~400 cells imaged in 7 independent data sets. RNA species were then clustered based on these average correlation coefficients using the same hierarchical clustering algorithm described above. Because of the large number of cells used for the analysis, it was found that the coarse spatial binning (2x2 regions per cell) was sufficient to capture the spatial correlation between genes and finer binning did not produce more significantly correlated groups.

To measure the distances of genes from the nuclei and from the cell edge, brightness thresholds on the cell images were first used to segment the nuclei and the cell edges identified. The distance from every RNA molecule to the nearest part of the nucleus and nearest part of the cell edge was then determined. For each data set, the average distance for each RNA species averaged over all the cells measured was then computed. These distances were averaged for the group I genes, group II genes or all genes. Only those RNA species with at least 10 counts per cell were used in this analysis to minimize statistical error on the distance values.

Gene ontology (GO) analysis. Groups of genes were selected from the hierarchical trees as discussed above. A collection of GO terms was determined for all measured RNA species as well as the RNA species associated with each group from the most recent human GO annotations using both the annotated GO terms and terms immediately upstream or downstream of the found annotations. The enrichment of these annotations was calculated from the ratio of the fraction of genes within each group that have this term to the fraction of all measured genes that have this term and the p-value for this enrichment was calculated via the hypergeometric function. Only statistically significantly enriched GO terms with a p-value less than 0.05 were considered.

While several embodiments of the present invention have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the functions and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the present invention. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the teachings of the present invention is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, the invention may be practiced otherwise than as specifically described and claimed. The present invention is directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles,

materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the scope of the present invention.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

The indefinite articles "a" and "an," as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean "at least one."

The phrase "and/or," as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with "and/or" should be construed in the same fashion, i.e., "one or more" of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the "and/or" clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to "A and/or B", when used in conjunction with open-ended language such as "comprising" can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, "or" should be understood to have the same meaning as "and/or" as defined above. For example, when separating items in a list, "or" or "and/or" shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as "only one of" or "exactly one of," or, when used in the claims, "consisting of," will refer to the inclusion of exactly one element of a number or list of elements. In general, the term "or" as used herein shall only be interpreted as indicating exclusive alternatives (i.e. "one or the other but not both") when preceded by terms of exclusivity, such as "either," "one of," "only one of," or "exactly one of." "Consisting essentially of," when used in the claims, shall have its ordinary meaning as used in the field of patent law.

As used herein in the specification and in the claims, the phrase "at least one," in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also

allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase "at least one" refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, "at least one of A and B" (or, equivalently, "at least one of A or B," or, equivalently "at least one of A and/or B") can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

When the word "about" is used herein in reference to a number, it should be understood that still another embodiment of the invention includes that number not modified by the presence of the word "about."

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

In the claims, as well as in the specification above, all transitional phrases such as "comprising," "including," "carrying," "having," "containing," "involving," "holding," "composed of," and the like are to be understood to be open-ended, i.e., to mean including but not limited to. Only the transitional phrases "consisting of" and "consisting essentially of" shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/329,683	01/27/2017	Xiaowei Zhuang	H0498.70509US02	7428

86110 7590 07/26/2021
Wolf, Greenfield & Sacks, P.C. (Harvard)
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER
LU, FRANK WEI MIN

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
07/26/2021	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

Patents_eOfficeAction@WolfGreenfield.com
H0498_eOfficeAction@WolfGreenfield.com

Response to Rule 312 Communication	Application No. 15/329,683	Applicant(s)
	Examiner	Art Unit

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address –

1. ☒ The amendment filed on *13 July 2021* under 37 CFR 1.312 has been considered, and has been:

a) ☒ entered.

b) ☐ entered as directed to matters of form not affecting the scope of the invention.

c) ☐ disapproved because the amendment was filed after the payment of the issue fee.
Any amendment filed after the date the issue fee is paid must be accompanied by a petition under 37 CFR 1.313(c)(1) and the required fee to withdraw the application from issue.

d) ☐ disapproved. See explanation below.

e) ☐ entered in part. See explanation below.

/C. Garner/

Publishing Division

Office of Data Management

Docket No.: H0498.70509US02
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	Xiaowei Zhuang
Application No.:	15/329,683
Confirmation No.:	7428
Filed:	January 27, 2017
For:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
Examiner:	F. W. Lu
Art Unit:	1634

Certificate of Electronic Filing under 37 C.F.R. § 1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: July 26, 2021 Electronic Signature for : /Trish McDonald/

NOTIFICATION OF LOSS OF ENTITLEMENT TO SMALL ENTITY STATUS

Mail Stop Issue Fee
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

Assignee, President and Fellows of Harvard College, hereby notifies the Patent and Trademark Office that it is not entitled to status as a small entity, and that the claim for small entity status is hereby withdrawn.

In compliance with 37 C.F.R. §§ 1.28(c)(1) and (c)(2), and out of an abundance of caution, the deficiency owed in connection with the fees paid as a small entity is itemized in the table below.

Date Paid	Fee Description (Fee Code)	Current Fee for Large Entity	Amount Previously Paid	Total Deficiency Owed
7/13/2021	2501 – Part B - Issue Fee	$1,200.00	$600.00	$600.00
			Total Deficiency Due:	**$600.00**

9421317.1

This paper also serves as a notification of loss of small entity status under 37 C.F.R. § 1.27(g)(2). It is requested that this error be excused as this filing is in compliance with the rules.

A Corrected Application Data Sheet is also being submitted herewith to update the entity size.

Please charge Deposit Account No. 23/2825 in the amount of $600.00, under Docket No. H0498.70509US02, from which the undersigned is authorized to draw.

Dated: July 26, 2021

Respectfully submitted,

Electronic signature: /Tani Chen/
Tani Chen, Sc.D.
Registration No.: 52,728
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

Electronic Acknowledgement Receipt

EFS ID:	43347879
Application Number:	15329683
International Application Number:	
Confirmation Number:	7428
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Customer Number:	86110
Filer:	Tani Chen/Patricia McDonald
Filer Authorized By:	Tani Chen
Attorney Docket Number:	H0498.70509US02
Receipt Date:	26-JUL-2021
Filing Date:	27-JAN-2017
Time Stamp:	19:39:24
Application Type:	U.S. National Stage under 35 USC 371

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	H049870509US02-TRN-TC.pdf	26176 03f9187b597a9251417091ea174645dfc49d8909	no	1
Warnings:					

Information:					
2	Notification of loss of entitlement to small entity status	H049870509US02-NLSES-TC.pdf	30995 d82a48ba846f51ae791b97352aa1fb55bdaccc9f	no	2
Warnings:					
Information:					
3	Application Data Sheet	H049870509US02-CADS-TC.pdf	66291 47d7d7deb131c327b45a2c27592c8a760a4997bd	no	8
Warnings:					
Information:					
This is not an USPTO supplied ADS fillable form					
			Total Files Size (in bytes):	123462	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

<u>New Applications Under 35 U.S.C. 111</u>
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.
<u>National Stage of an International Application under 35 U.S.C. 371</u>
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.
<u>New International Application Filed with the USPTO as a Receiving Office</u>
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/329,683-Conf. #7428
Filing Date	January 27, 2017
First Named Inventor	Xiaowei Zhuang
Art Unit	1634
Examiner Name	F. W. Lu
Attorney Docket Number	H0498.70509US02

ENCLOSURES (*Check all that apply*)

- [ ] Fee Transmittal Form
  - [ ] Fee Attached
- [ ] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [ ] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) (please Identify below):
  - Notification of Loss of Small Entity Status
  - Corrected Application Data Sheet

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Tani Chen/		
Printed name	Tani Chen, Sc.D.		
Date	July 26, 2021	Reg. No.	52,728

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: July 26, 2021 Electronic Signature for : /Trish McDonald/

9421315.1

Corrected Application Data Sheet

Inventor Information

Inventor Number::	1
Given Name::	Xiaowei
Family Name::	Zhuang
City of Residence::	~~Lexington~~Cambridge
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	~~5 Luongo Farm Lane~~ c/o Harvard University Smith Center - Suite 727; 1350 Massachusetts Ave.
City of mailing address::	~~Lexington~~ Cambridge
State or Province of mailing address::	MA
Country of mailing address::	US
Postal or Zip Code of mailing address::	~~02421~~ 02138

Inventor Number::	2
Given Name::	Kok-Hao
Family Name::	Chen
City of Residence::	~~Singapore~~Cambridge
State or Province of Residence:	MA
Country of Residence::	~~Republic of Singapore~~US

Street of mailing address::	~~37 Jalan Lepas~~ c/o Harvard University Smith Center - Suite 727; 1350 Massachusetts Ave.
City of mailing address::	~~Singapore~~ Cambridge
State or Province of mailing address::	MA
Country of mailing address::	~~Republic of Singapore~~ US
Postal or Zip Code of mailing address::	~~537961~~ 02138

Inventor Number::	3
Given Name::	Alistair
Family Name::	Boettiger
City of Residence::	Cambridge
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	~~1550 Cambridge Street~~ c/o Harvard University ~~Apt. #2~~ Smith Center - Suite 727; 1350 Massachusetts Ave.
City of mailing address::	Cambridge
State or Province of mailing address::	MA
Country of mailing address::	US
Postal or Zip Code of mailing address::	~~02139~~ 02138

Inventor Number::	4
Given Name::	Jeffrey
Middle Name::	R.
Family Name::	Moffitt
City of Residence::	~~Somerville~~ Cambridge
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	~~175 Beacon Street~~ c/o Harvard University ~~Apt. #108~~ Smith Center - Suite 727; 1350 Massachusetts Ave.
City of mailing address::	~~Somerville~~ Cambridge
State or Province of mailing address::	MA
Country of mailing address::	US
Postal or Zip Code of mailing address::	~~02143~~ 02138

Inventor Number::	5
Given Name::	Siyuan
Family Name::	Wang
City of Residence::	Cambridge
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	~~28 Fernald Drive~~ c/o Harvard University

	~~Apt. #31~~ Smith Center - Suite 727; 1350 Massachusetts Ave.
City of mailing address::	Cambridge
State or Province of mailing address::	MA
Country of mailing address::	US
Postal or Zip Code of mailing address::	02138

Correspondence Information

Correspondence Customer Number::	86110

Application Information

Application Number::	15/329,683
Filing Date::	01/27/17
Application Type::	~~Regular~~ Nonprovisional
Subject Matter::	Utility
CD-ROM or CD-R?::	None
Sequence submission?::	None
Computer Readable Form (CRF)?::	No
Title::	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
Attorney Docket Number::	H0498.70509US02
Request for Early Publication?::	No
Request for Non-Publication?::	No
Small Entity?::	~~Yes~~No

Petition included?::	No
Licensed US Govt. Agency::	National Institutes of Health
Contract or Grant Numbers::	GM096450
Portions or all of the application associated with this Application Data Sheet may fall under a Secrecy Order pursuant to 37 CFR 5.2::	No

Representative Information

Representative Customer Number::	86110

Domestic Priority Information

Application::	Continuity Type::	Parent Application::	Parent Filing Date::	Prior Appl Status::
This Application	National Stage of	PCT/US2015/042556	07/29/15	
PCT/US2015/042556	An application claiming the benefit under 35 USC 119(e)	62/142653	04/03/15	
PCT/US2015/042556	An application claiming the benefit under 35 USC 119(e)	62/050636	09/15/14	
PCT/US2015/042556	An application claiming the benefit under 35 USC 119(e)	62/031062	07/30/14	

Foreign Priority Information

Applicant Information

Applicant Number::	1
Applicant Type::	Assignee
Organization Name::	President and Fellows of Harvard College
Street of mailing address::	17 Quincy Street
City of mailing address::	Cambridge
State or Province of mailing address::	MA
Country of mailing address::	US
Postal or Zip Code of mailing address::	02138

Assignee Information Including Non-Applicant Assignee Information

Assignee Number::	1
Organization Name::	President and Fellows of Harvard College
Street of mailing address::	17 Quincy Street
City of mailing address::	Cambridge
State or Province of mailing address::	MA
Country of mailing address::	US
Postal or Zip Code of mailing address::	02138

Signature:

NOTE: This Application Data Sheet must be signed in accordance with 37 CFR 1.33(b). **However, if this Application Data Sheet is submitted with the INITIAL filing of the application and either box A or B is not checked in subsection 2 of the "Authorization or Opt-Out of Authorization to Permit Access" section, then this form must also be signed in accordance with 37 CFR 1.14(c).**

This Application Data Sheet **must** be signed by a patent practitioner if one or more of the applicants is a **juristic entity** (e.g., corporation or association). If the applicant is two or more joint inventors, this form must be signed by a patent practitioner, **all** joint inventors who are the applicant, or one or more joint inventor-applicants who have been given power of attorney (e.g., see USPTO Form PTO/AIA/81) on behalf of **all** joint inventor-applicants.

See 37 CFR 1.4(d) for the manner of making signatures and certifications.

Signature	/Tani Chen/	Date (YYYY-MM-DD)	2021-07-26
Name	Tani Chen, Sc.D.	Registration Number	52,728

PTO/SB/92 (07-16)
Approved for use through 12/31/2020. OMB 0561-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Certificate of Mailing or Transmission under 37 CFR 1.8

I hereby certify that this correspondence is being:

1. ☐ Deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to:

   Commissioner for Patents
   P.O. Box 1450
   Alexandria, VA 22313-1450

   on _____ .
   Date

2. ☐ Facsimile transmitted to the United States Patent and Trademark Office, or

   OR

3. ☒ EFS-Web transmitted to the USPTO.

/Trish McDonald/	July 26, 2021
Signature	Date
Trish McDonald	617.646.8000
Typed or printed name	Telephone number

Note: Each paper must have its own certificate of mailing or transmission, or this certificate must identify each submitted paper:

Corrected Application Data Sheet (7 pages)

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	ISSUE DATE	PATENT NO.	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/329,683	08/24/2021	11098303	H0498.70509US02	7428

86110 7590 08/04/2021

Wolf, Greenfield & Sacks, P.C. (Harvard)
600 Atlantic Avenue
Boston, MA 02210-2206

ISSUE NOTIFICATION

The projected patent number and issue date are specified above.

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)
(application filed on or after May 29, 2000)

The Patent Term Adjustment is 213 day(s). Any patent to issue from the above-identified application will include an indication of the adjustment on the front page.

If a Continued Prosecution Application (CPA) was filed in the above-identified application, the filing date that determines Patent Term Adjustment is the filing date of the most recent CPA.

Applicant will be able to obtain more detailed information by accessing the Patent Application Information Retrieval (PAIR) WEB site (http://pair.uspto.gov).

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Application Assistance Unit (AAU) of the Office of Data Management (ODM) at (571)-272-4200.

APPLICANT(s) (Please see PAIR WEB site http://pair.uspto.gov for additional applicants):

Xiaowei Zhuang, Lexington, MA;
President and Fellows of Harvard College, Cambridge, MA;
Kok-Hao Chen, Singapore, SINGAPORE;
Alistair Boettiger, Cambridge, MA;
Jeffrey R. Moffitt, Somerville, MA;
Siyuan Wang, Cambridge, MA;

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The USA offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to encourage and facilitate business investment. To learn more about why the USA is the best country in the world to develop technology, manufacture products, and grow your business, visit SelectUSA.gov.

IR103 (Rev. 10/09)

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY.DOCKET NO./TITLE	REQUEST ID
15/329,683	01/27/2017	Xiaowei Zhuang	H0498.70509US02	145667

Acknowledgement of Loss of Entitlement to Entity Status Discount

The entity status change request below filed through Private PAIR on 08/17/2021 has been accepted.

CERTIFICATIONS:

Change of Entity Status:

X Applicant changing to regular undiscounted fee status.

NOTE: Checking this box will be taken to be notification of loss of entitlement to small or micro entity status, as applicable.

This portion must be completed by the signatory or signatories making the entity status change in accordance with 37 CFR 1.4(d)(4).

Signature:	/Tani Chen/
Name:	Tani Chen
Registration Number:	52728

Docket No.: H0498.70509US02
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	Xiaowei Zhuang
Application No.:	15/329,683
Confirmation No.:	7428
Filed:	January 27, 2017
Patent No.:	11,098,303
Issued:	August 24, 2021
For:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
Examiner:	F.W. Lu
Art Unit:	1634

Certificate of Electronic Filing under 37 C.F.R. § 1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: October 22, 2021 Signature: /Rachel Moulton/ (Rachel Moulton)

APPLICATION FOR PATENT TERM ADJUSTMENT UNDER 37 C.F.R. § 1.705(b)

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Commissioner:

The present Application ("Request") for Patent Term Adjustment ("PTA") is being submitted pursuant to 37 C.F.R. § 1.705(b).

I. SUMMARY

The Office has determined that the above-referenced Patent ("Patent") is entitled to 213 days of PTA. Applicant respectfully disagrees with the Office's determination of PTA. The Office's determination of PTA was incorrect because the Office miscalculated the reductions of period of adjustment under 35 U.S.C. § 154(b)(2)(C) ("applicant delays"). The Office determined the applicant delays to be 217 days. Applicant submits that the applicant delays should be 111 days, not

217 days. As a result, Applicant submits that the PTA should be ***319 days***, not 213 days. Applicant respectfully requests that the Office reconsider its determination of PTA.

A statement of facts and description of the correct PTA determination are provided below.

II. STATEMENT OF FACTS

1. The Patent was issued on August 24, 2021, from U.S. Patent Application Number 15/329,683 ("'683 application").

2. The '683 application was filed on January 27, 2017, and is a national stage entry under § 371 of International PCT Application Number PCT/US2015/042556 ("'556 application").

3. The '556 application was filed on July 29, 2015, and claims priority under § 119(e) to U.S. Provisional Application Numbers 62/142,653, filed 04/03/15, 62/050,636, filed 09/15/14, and 62/031,062, filed 07/30/14.

4. In the Transmittal Letter submitted with the filing of the '683 application, the check box "This is an express request to begin national examination procedures (35 U.S.C. 371 (f))" is not checked.

5. On October 22, 2020, the Office mailed a non-final Office Action ("OA").

6. On January 22, 2021, Applicant filed a Response to the OA ("Response to OA").

7. On February 18, 2021, Frank Lu, the Examiner; Jeffrey Moffitt, an inventor; and Tani Chen, an Applicant's representative, held a Telephonic Interview.

8. On February 24, 2021, the Office mailed an Applicant-Initiated Interview Summary ("Interview Summary"). In the Interview Summary, the Office stated that "Dr. [M]offit, Mr. Chen,

and the examiner discussed the inventions and possible amendment. The interview helped the examiner to better understand the inventions of applicants."

9. On March 26, 2021, Applicant filed a Supplemental Response to the OA ("Supplemental Response").

10. On May 26, 2021, the Office mailed a Notice of Allowance ("NOA").

11. On June 4, 2021, the Office mailed a Notice to File Corrected Application Papers ("Notice"). In the Notice, the Office stated that "[i]n response to this notice, the applicant must submit a substitute specification in which the pages are so numbered."

12. On July 13, 2021, Applicant paid the Issue Fee, under the small-entity status, and filed a Substitute Specification.

13. On July 26, 2021, the Office mailed a Response to Rule 312 Communication.

14. On July 26, 2021, Applicant filed a Corrected Application Data Sheet ("CADS") and a Notification of Loss of Entitlement to Small Entity Status ("Notification"). In the CADS and Notification, Applicant withdrew the claim for small entity status. Also in the CADS, Applicant changed the mailing addresses of all five inventors.

15. No part of the term of the Patent has been disclaimed.

III. ANALYSIS

A. Legal Standard

1. 35 U.S.C. § 154(b)(2)(C)(i) provides:

> The period of adjustment of the term of a patent under paragraph (1) shall be reduced by a period equal to the period of time during which the applicant failed to engage in reasonable efforts to conclude prosecution of the application.

2. 37 C.F.R. § 1.704(c) as revised by 85 Federal Register 36335, June 16, 2020, has been effective since July 16, 2020.

3. § 1.704(c)(8) effective July 16, 2020 provides:

> (c) Circumstances that constitute a failure of the applicant to engage in reasonable efforts to conclude processing or examination of an application also include the following circumstances, which will result in the following reduction of the period of adjustment set forth in § 1.703 to the extent that the periods are not overlapping: . . .
>
> (8) Submission of a supplemental reply or other paper, other than a supplemental reply or other paper expressly requested by the examiner, after a reply has been filed, in which case the period of adjustment set forth in § 1.703 shall be reduced by the number of days, if any, beginning on the day after the date the initial reply was filed and ending on the date that the supplemental reply or other such paper was filed . . . .

4. § 1.704(c)(10) effective July 16, 2020 provides:

> (c) Circumstances that constitute a failure of the applicant to engage in reasonable efforts to conclude processing or examination of an application also include the following circumstances, which will result in the following reduction of the period of adjustment set forth in § 1.703 to the extent that the periods are not overlapping: . . .
>
> (10) Submission of an amendment under §1.312 or other paper, other than an amendment under§ 1.312 or other paper expressly requested by the Office or a request for continued examination incompliance with § 1.114, after a notice of allowance has been given or mailed, in which case the period of adjustment set forth in § 1.703 shall be reduced by the number of days, if any, beginning on the day after the date of mailing of the notice of allowance under 35 U.S.C. 151 and ending on the date the amendment under § 1.312 or other paper was filed . . . .

5. § 1.704(c)(10) effective July 16, 2020 includes changes applicable only to original utility and plant patents issuing from applications filed on or after May 29, 2000, in which a notice of allowance was mailed on or after July 16, 2020 or in which a notice of allowance was mailed prior to July 16, 2020 and patentee has requested applicability in a timely filed request for reconsideration under § 1.705.

6. MPEP § 2732 provides, in relevant part:

> The submission of the following papers after a "Notice of Allowance" is not considered a failure to engage in reasonable efforts to conclude processing or examination of an application: . . . (4) change of address; (5) change of status (micro/small/not small entity status) . . . ."

B. Determination of PTA

1. A Delays

The examination delays under 35 U.S.C. § 154(b)(1)(A) as determined by the Office ("A delays") are 263 days. Applicant does not dispute the Office's determination of A delays.

2. B Delays

The examination delays under 35 U.S.C. § 154(b)(1)(B) as determined by the Office ("B delays") are 167 days. Applicant does not dispute the Office's determination of B delays.

3. C Delays

The examination delays under 35 U.S.C. § 154(b)(1)(C) as determined by the Office ("C delays") are 0 days. Applicant does not dispute the Office's determination of C delays.

4. Non-Overlapping Office Delays

The non-overlapping Office delays as determined by the Office are 430 days. Applicant does not dispute the Office's determination of the non-overlapping Office delays.

5. Applicant Delays

(a) Substitute Specification

(i) The number of days beginning on the day after the date the Substitute Specification was filed and ending on the date the Response to Rule 312 Communication was mailed is 14 days.

(ii) The Office appeared to have determined that the 14 days provided in Section III(B)(5)(a)(i) of this Request is an applicant delay under § 1.704(c)(10).

(iii) Applicant respectfully disagrees with the Office's determination of the applicant delay provided in Section III(B)(5)(a)(ii) of this Request.

(iv) The Patent is an original utility patent issuing from an application filed after May 29, 2000, in which a notice of allowance was mailed after July 16, 2020. Therefore, § 1.704(c)(10) effective July 16, 2020 is applicable in the determination of the applicant delay caused by filing the Substitute Specification, or the lack thereof.

(v) In the Notice, the Office stated that "[i]n response to this notice, the applicant must submit a substitute specification in which the pages are so numbered." Applicant filed the Substitute Specification for the sole purpose of responding to the Notice. Therefore, the Substitute Specification is "an amendment under§ 1.312 . . . expressly requested by the Office" under § 1.704(c)(10) effective July 16, 2020. Therefore, Applicant's filing of the Substitute Specification is not an applicant delay under § 1.704(c)(10) effective July 16, 2020.

(vi) Applicant's filing of the Substitute Specification is not an applicant delay under 35 U.S.C. § 154(b)(2)(C) or 37 C.F.R. § 1.704(c).

(vii) The Office's determination that the 14 days provided in Section III(B)(5)(a)(ii) of this Request is an applicant delay was incorrect. These 14 days should have been removed from applicant delays.

(b) CADS and Notification

(i) The number of days beginning on the day after the date the CADS and Notification were filed and ending on the date the Patent was issued is 29 days.

(ii) The Office appeared to have determined that the 29 days provided in Section III(B)(5)(b)(i) of this Request is an applicant delay under § 1.704(c)(10).

(iii) Applicant respectfully disagrees with the Office's determination of the applicant delay provided in Section III(B)(5)(b)(ii) of this Request.

(iv) The Patent is an original utility patent issuing from an application filed after May 29, 2000, in which a notice of allowance was mailed after July 16, 2020. Therefore, § 1.704(c)(10) effective July 16, 2020 is applicable in the determination of the applicant delay caused by filing the CADS and Notification, or the lack thereof.

(v) In the CADS and Notification, Applicant withdrew the claim for small entity status. Also in the CADS, Applicant changed the mailing addresses of all five inventors. In the CADS and Notification, Applicant made no other changes to the '683 application. § 1.704(c)(10) effective July 16, 2020 does not provide whether a submission after a Notice of Allowance to change address or to change entity status is or is not an applicant delay. However, under MPEP § 2732, Applicant's filing of the CADS and Notification was a "submission of the following papers after a 'Notice of Allowance'[:] . . . (4) change of address; [and] (5) change of status (micro/small/not

small entity status) . . . ." Therefore, Applicant's filing of the CADS and Notification is not an applicant delay under § 1.704(c)(10) effective July 16, 2020.

(vi) Applicant's filing of the CADS and Notification is not an applicant delay under 35 U.S.C. § 154(b)(2)(C) or 37 C.F.R. § 1.704(c).

(vii) The Office's determination that the 29 days provided in Section III(B)(5)(b)(ii) of this Request is an applicant delay was incorrect. These 29 days should have been removed from applicant delays.

(c) Supplemental Response

(i) The number of days beginning on the day after the date the Response to OA was filed and ending on the date that the Supplemental Response was filed is 63 days.

(ii) The Office appeared to have determined that the 63 days provided in Section III(B)(5)(c)(i) of this Request is an applicant delay under § 1.704(c)(8).

(iii) Applicant respectfully disagrees with the Office's determination of the applicant delay provided in Section III(B)(5)(c)(ii) of this Request.

(iv) The Patent is an original utility patent issuing from an application filed after May 29, 2000, in which a notice of allowance was mailed after July 16, 2020. Therefore, § 1.704(c)(8) effective July 16, 2020 is applicable in the determination of the applicant delay caused by filing the Supplemental Response, or the lack thereof.

(v) In the Interview Summary, the Office stated that "Dr. [M]offit, Mr. Chen, and the examiner discussed the inventions and possible amendment. The interview helped the examiner to better understand the inventions of applicants." Applicant filed the Supplemental Response for the sole purpose of addressing the subject matter discussed during the telephonic Interview. Therefore, the

Supplemental Response is "a supplemental reply . . . expressly requested by the examiner" under § 1.704(c)(8) effective July 16, 2020. Therefore, Applicant's filing of the Supplemental Response is not an applicant delay under §1.704(c)(8) effective July 16, 2020.

(vi) Applicant's filing of the Supplemental Response is not an applicant delay under 35 U.S.C. § 154(b)(2)(C) or 37 C.F.R. § 1.704(c).

(vii) The Office's determination that the 63 days provided in Section III(B)(5)(c)(ii) of this Request is an applicant delay was incorrect. These 63 days should have been removed from applicant delays.

(d) Additional Applicant Delays

The Office appeared to have determined additional applicant delays to be 111 days (24 days + 87 days). Applicant does not dispute the Office's determination of these additional applicant delays.

(e) Applicant determines the total applicant delays to be 111 days (24 days + 87 days).

6. Applicant determines the PTA to be 319 days (430 days – 111 days).

IV. CONCLUSION

The Office's determination of PTA (213 days) was incorrect because the Office miscalculated the applicant delays.

First, the Office determined as an applicant delay the period beginning on the day after the date the Substitute Specification was filed and ending on the date the Response to Rule 312 Communication was mailed. This period should not have been determined as an applicant delay because the Substitute Specification is "an amendment under § 1.312 . . . expressly requested by the Office" under § 1.704(c)(10) effective July 16, 2020.

Second, the Office determined as an applicant delay the period beginning on the day after the date the CADS and Notification were filed and ending on the date the Patent was issued. This period should not have been determined as an applicant delay because Applicant's filing of the CADS and Notification was a "submission of the following papers after a 'Notice of Allowance'[:] . . . (4) change of address; [and] (5) change of status (micro/small/not small entity status) . . ." under MPEP § 2732.

Third, the Office determined as an applicant delay the period beginning on the day after the date the Response to OA was filed and ending on the date that the Supplemental Response was filed. This period should not have been determined as an applicant delay because the Supplemental Response is "a supplemental reply . . . expressly requested by the examiner" under § 1.704(c)(8) effective July 16, 2020.

Applicant respectfully requests that the Office reconsider its determination of PTA and provide Applicant with ***319 days*** of PTA.

The deadline for filing an Application for PTA under 37 C.F.R. § 1.705(b) is October 24, 2021. Applicant submits that the Request is timely submitted.

Applicant submits herewith the fee of $210.00 for filing an Application for PTA as set forth in § 1.18(e). Please charge any fee deficiency, or credit any overpayment, to our Deposit Account No. 23/2825, under Docket No. H0498.70509US02, from which the undersigned is authorized to draw.

Dated: October 22, 2021

Respectfully submitted,

By /Tani Chen/
Tani Chen, Sc.D.
Registration No.: 52,728
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

9720759.1

FEE TRANSMITTAL

Complete if known	
Application Number	Patent #: 11,098,303
Filing Date	Issued: August 24, 2021
First Named Inventor	Xiaowei Zhuang
Examiner Name	F. W. Lu
Art Unit	1634
Practitioner Docket No.	H0498.70509US02

☐ Applicant asserts small entity status. See 37 CFR 1.27.

☐ Applicant certifies micro entity status. See 37 CFR 1.29. Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

TOTAL AMOUNT OF PAYMENT ($)210.00

METHOD OF PAYMENT (check all that apply)

☐ Check ☒ Credit Card ☐ Money Order ☐ None ☐ Other (please identify): _____

☒ Deposit Account Deposit Account Number: 23/2825 Deposit Account Name: Wolf, Greenfield & Sacks, P.C.

For the above-identified deposit account, the Director is hereby authorized to (check all that apply):

☒ Charge fee(s) indicated below

☐ Charge fee(s) indicated below, **except for the filing fee**

☒ Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17, **except for any excess claims fees or multiple dependent claim fee**

☐ Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17

☒ Credit any overpayment of fee(s)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES (U = undiscounted fee; S = small entity fee; M = micro entity fee)

	FILING FEES			SEARCH FEES			EXAMINATION FEES			
Application Type	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**Fees Paid ($)**
Utility	320	160*	80	700	350	175	800	400	200	
Design	220	110	55	160	80	40	640	320	160	
Plant	220	110	55	440	220	110	660	330	165	
Reissue	320	160	80	700	350	175	2,320	1,160	580	
Provisional	300	150	75	0	0	0	0	0	0	

* The $160 small entity status filing fee for a utility application is further reduced to $80 for a small entity status applicant who files the application via EFS-Web.

2. EXCESS CLAIM FEES

Fee Description	**Undiscounted Fee ($)**	**Small Entity Fee ($)**	**Micro Entity Fee ($)**
Each claim over 20 (including Reissues)	100	50	25
Each independent claim over 3 (including Reissues)	480	240	120
Multiple dependent claims	860	430	215

Total Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**
_____ - 20 or HP =	_____ x	_____ =	_____

HP = highest number of total claims paid for, if greater than 20.

Multiple Dependent Claims	
Fee ($)	**Fee Paid ($)**
_____	_____

Indep. Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**
_____ - 3 or HP =	_____ x	_____ =	_____

HP = highest number of independent claims paid for, if greater than 3.

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper (excluding electronically filed sequence or computer listings under 37 CFR 1.52(e)), the application size fee due is $420 ($210 for small entity) ($105 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets	**Extra Sheets**	**Number of each additional 50 or fraction thereof**	**Fee ($)**	**Fee Paid ($)**
_____ - 100 =	_____ /50 =	_____ (round **up** to a whole number) x	_____ =	_____

4. OTHER FEE(S)

	Fees Paid ($)
Non-English specification, $130 fee (no small or micro entity discount)	
Non-electronic filing fee under 37 CFR 1.16(t) for a utility application, $400 fee ($200 small or micro entity)	
Other (e.g., late filing surcharge): 1455 Filing an application for patent term adjustment	210.00

SUBMITTED BY

Signature	/Tani Chen/	Registration No. (Attorney/Agent)	52,728	Telephone	617.646.8000
Name (Print/Type)	Tani Chen, Sc.D.			Date	October 22, 2021

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: October 22, 2021 Signature: /Rachel Moulton/ (Rachel Moulton)

9721645.1

Electronic Patent Application Fee Transmittal

Application Number:	15329683
Filing Date:	27-Jan-2017
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Filer:	Tani Chen/Rachel Moulton
Attorney Docket Number:	H0498.70509US02

Filed as Large Entity

Filing Fees for U.S. National Stage under 35 USC 371

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
APPLICATION FOR PATENT TERM ADJUSTMENT	1455	1	210	210
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension-of-Time:				
Miscellaneous:				
			Total in USD ($)	**210**

Electronic Acknowledgement Receipt

EFS ID:	44104835
Application Number:	15329683
International Application Number:	
Confirmation Number:	7428
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Customer Number:	86110
Filer:	Tani Chen/Rachel Moulton
Filer Authorized By:	Tani Chen
Attorney Docket Number:	H0498.70509US02
Receipt Date:	22-OCT-2021
Filing Date:	27-JAN-2017
Time Stamp:	17:38:39
Application Type:	U.S. National Stage under 35 USC 371

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$210
RAM confirmation Number	E20210LH40001470
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Transmittal Letter	H049870509US02-TRN-TC.pdf	26265 01d4f1cda1158f2d37d2828662a108289c3125ca	no	1
Warnings:					
Information:					
2	Request for Certificate of Correction	H049870509US02-RCOC-TC.pdf	25962 ec98ecd6c945a061865c81068a38e418c9e57a3c	no	2
Warnings:					
Information:					
3	Request for Certificate of Correction	H049870509US02-COC-TC.pdf	17307 82d3e5356b94dd4b28ed3830c5f4d8dc0bd7534e	no	1
Warnings:					
Information:					
4	Patent Term Adjustment Petition	H049870509US02-RPTA-TC.pdf	63191 905f43812a6f619e847830bd101dfb95020b4460	no	10
Warnings:					
Information:					
5	Fee Worksheet (SB06)	H049870509US02-FEE-TC.pdf	32804 91a758223c7e39c7ad1595204565dc86c5ef53de	no	1
Warnings:					
Information:					
6	Fee Worksheet (SB06)	fee-info.pdf	38124 2d53f53c7e6e761d3ecf12d9738daf12725b1d4c	no	2
Warnings:					
Information:					

Total Files Size (in bytes):	203653

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	Patent#: 11,098,303
Filing Date	Issued: August 24, 2021
First Named Inventor	Xiaowei Zhuang
Art Unit	1634
Examiner Name	F. W. Lu
Attorney Docket Number	H0498.70509US02

ENCLOSURES (*Check all that apply*)

- [x] Fee Transmittal Form
  - [ ] Fee Attached
- [ ] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [ ] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) (please Identify below):
  - Request for Certificate of Correction
  - Certificate of Correction
  - Application for Patent Term Adjustment

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Tani Chen/		
Printed name	Tani Chen, Sc.D.		
Date	October 22, 2021	Reg. No.	52,728

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: October 22, 2021 Signature: /Rachel Moulton/ (Rachel Moulton)

9700599.1

Docket No.: H0498.70509US02
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	Xiaowei Zhuang
Application No.:	15/329,683
Confirmation No.:	7428
Filed:	January 27, 2017
Patent No.:	11,098,303
Issued:	August 24, 2021
For:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
Examiner:	F. W. Lu
Art Unit:	1634

Certificate of Electronic Filing under 37 C.F.R. § 1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: October 22, 2021 Signature: /Rachel Moulton/ (Rachel Moulton)

REQUEST FOR CERTIFICATE OF CORRECTION

Attention: Certificate of Correction Branch
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

Upon reviewing the above-identified patent, the undersigned noted typographical errors which should be corrected.

On the first page, item (72) should read:

(72) Inventors: Xiaowei Zhuang, Cambridge, MA (US);
Kok-Hao Chen, Cambridge, MA (US);
Alistair Boettiger, Cambridge, MA (US);
Jeffrey R. Moffitt, Cambridge, MA (US);
Siyuan Wang, Cambridge, MA (US)

In the Claims:

At column 85, claim 1, line 50 currently reads (the "ent" is the second half of the word "different," which was hyphenated across lines 49 and 50):

ent encoding nucleic acid probes

At column 85, claim 1, line 50 should read (additions underlined):

ent encoding nucleic acid probes located in its different locations;

This request does not involve correction of Applicant errors. Accordingly, no fee is required.

Transmitted herewith is a proposed Certificate of Correction effecting such amendment. The undersigned respectfully solicits the granting of the requested Certificate of Correction.

No fee is believed due with this Request for Certificate of Correction. However, if a fee is due, please charge Deposit Account No. 23/2825, under Docket No. H0498.70509US02 from which the undersigned is authorized to draw.

Dated: October 22, 2021

Respectfully submitted,

By /Tani Chen/

Tani Chen, Sc.D.
Registration No.: 52,728
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

9700601.1

PTO/SB/44 (09-07)
Approved for use through 01/31/2020. OMB 0651-0033
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.
(Also Form PTO-1050)

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 11,098,303

APPLICATION NO. : 15/329,683

ISSUE DATE : August 24, 2021

INVENTOR(S) : Xiaowei Zhuang et al.

It is certified that errors appear in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

On the first page, item (72) should read:

(72) Inventors: Xiaowei Zhuang, Cambridge, MA (US);
Kok-Hao Chen, Cambridge, MA (US);
Alistair Boettiger, Cambridge, MA (US);
Jeffrey R. Moffitt, Cambridge, MA (US);
Siyuan Wang, Cambridge, MA (US)

In the Claims:

Column 85, claim 1, line 50 should read:

ent encoding nucleic acid probes located in its different locations;

MAILING ADDRESS OF SENDER (Please do not use Customer Number below):
Tani Chen, Sc.D.
WOLF, GREENFIELD & SACKS, P.C. 1
600 Atlantic Avenue
Boston, Massachusetts 02210-2206

9700600.1

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/329,683	01/27/2017	Xiaowei Zhuang	H0498.70509US02	7428

86110 7590 11/12/2021

Wolf, Greenfield & Sacks, P.C. (Harvard)
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER	
LU, FRANK WEI MIN	

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
11/12/2021	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

H0498_eOfficeAction@WolfGreenfield.com
Patents_eOfficeAction@WolfGreenfield.com

UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

Patent No.: 11098303
Issue Date: 08/24/2021
Appl. No.: 15/329,683
Filed: 01/27/2017

PART (A) RESPONSE FOR CERTIFICATES OF CORRECTION

This is a decision on the Certificate of Correction request filed 22 October 2021.

The request for issuance of Certificate of Correction for the above-identified correction(s) under the provisions of 37 CFR 1.322 and/or 1.323 is hereby:

(Check one)
☑ Approved ☐ Approved in Part ☐ Denied

Comments: ____

PART (B) PETITION UNDER 37 CFR 1.324 OR 37 CFR 1.48

☐ This is a decision on the petition filed ____ to correct inventorship under 37 CFR 1.324.

☐ This is a decision on the request under 37 CFR 1.48, petition filed ____. In view of the fact that the patent has already issued, the request under 37 CFR 1.48 has been treated as a petition to correct inventorship under 37 CFR 1.324.

The petition is hereby: ☐ Granted ☐ Dismissed

Comment: ____

The patented filed is being forwarded to Certificate of Corrections Branch for issuance of a certificate naming only the actual inventor or inventors.

/DAVE T NGUYEN/
Supervisory Patent Examiner, Art Unit 1634
Technology Center 1600
Phone: (571)272-0731

Certificates of Correction Branch email: CustomerServiceCoC@uspto.gov CoC Central Phone Number: (703) 756-1814

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 11,098,303 B2
APPLICATION NO. : 15/329683
DATED : August 24, 2021
INVENTOR(S) : Xiaowei Zhuang et al.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

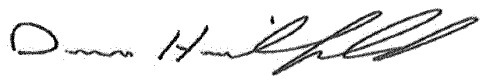
On the Title Page

Item (72) should read:
(72) Inventors: Xiaowei Zhuang, Cambridge, MA (US);
Kok-Hao Chen, Cambridge, MA (US);
Alistair Boettiger, Cambridge, MA (US);
Jeffrey R. Moffitt, Cambridge, MA (US);
Siyuan Wang, Cambridge, MA (US)

In the Claims

Column 85, Claim 1, Line 50 should read:
ent encoding nucleic acid probes located in its different locations;

Signed and Sealed this
Seventh Day of December, 2021



Drew Hirshfeld
Performing the Functions and Duties of the
Under Secretary of Commerce for Intellectual Property and
Director of the United States Patent and Trademark Office

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/329,683	01/27/2017	Xiaowei Zhuang	H0498.70509US02	7428

86110 7590 03/18/2022

Wolf, Greenfield & Sacks, P.C. (Harvard)
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER	
LU, FRANK WEI MIN	

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
03/18/2022	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

H0498_eOfficeAction@WolfGreenfield.com
Patents_eOfficeAction@WolfGreenfield.com

UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

In re Patent No. 11,098,303	:	
Issue Date: 24 Aug 2021	:	DECISION ON PETITION
Application No. 15/329,683	:	
Filing or 371(c) Date: 27 Jan 2017	:	
Attorney Docket No. H0498.70509US02	:	

This is a decision on Patentee's request for reconsideration pursuant to 37 C.F.R. § 1.705(b) filed October 22, 2021, requesting that the Office adjust the patent term adjustment from 213 days to 319 days.

The petition to correct the patent term adjustment indicated on the above-identified patent to indicate that the term of the above-identified patent is extended or adjusted by 319 days is **GRANTED**.

The Office acknowledges the receipt of the $210.00 fee set forth in 37 C.F.R. § 1.18(e). No additional fees are required. A review of the image file wrapper shows that this patent is not subject to a terminal disclaimer.

The application is being forwarded to the Certificates of Correction Branch for issuance of a certificate of correction. The Office will issue a certificate of correction indicating that the term of the above-identified patent is extended or adjusted by **319 days.**

Telephone inquiries specific to this matter should be directed to the undersigned at (571) 272-3225.

/Paul Shanoski/
Paul Shanoski
Attorney Advisor
Office of Petitions

Encl: Adjusted PTA calculation
DRAFT Certificate of Correction

DRAFT COPY

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT : 11,098,303 B2

DATED : **Aug. 24, 2021** **DRAFT**

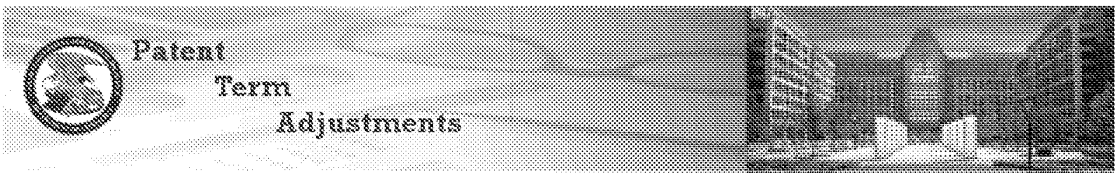
INVENTOR(S) : **Zhuang et al.**

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

On the cover page,

[*] Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 USC 154(b) by 213 days

Delete the phrase "by 213 days" and insert – by 319 days --



PTA/PTE Information

Application Number*: 15329683 Open Explanation of PTA Calculation Explanation of PTE Calculation

Print

PTA Calculations for Application: 15329683

Application Filing Date	01/27/2017	OverLapping Days Between (A and B) or (A and C)	0
Issue Date of Patent	08/24/2021	Non-Overlapping USPTO Delays	430
A Delays	263	PTO Manual Adjustment	106
B Delays	167	Applicant Delay (APPL)	217
C Delays	0	Total PTA (days)	319

* - Sorted Column

File Contents History

Action Number	Action Recorded Date	Action Due Date	Action Code	Action Description	Duration PTO	Duration APPL	Parent Action Number
147	03/17/2022		P028	Adjustment of PTA Calculation by PTO	29		0
146	03/17/2022		P028	Adjustment of PTA Calculation by PTO	14		0
145	03/17/2022		P028	Adjustment of PTA Calculation by PTO	63		0
133.5	08/24/2021	01/30/2020	PTA36M	PTA 36 Months	187		133.4
133	08/24/2021		PTAC	Patent Issue Date Used in PTA Calculation			0
132	07/30/2021		EFDC	Export to Final Data Capture			0
131	07/29/2021		FIDC	Finished Initial Data Capture			0
130	07/29/2021		D1935	Dispatch to FDC			0
129	07/29/2021		PILS	Application Is Considered Ready for Issue			0
128	07/26/2021	07/26/2021	LET.	Miscellaneous Incoming Letter		29	133
126	07/26/2021		EML_NTR	Email Notification			0
125	07/26/2021	07/13/2021	MN271	Mail Response to 312 Amendment (PTO-271)		14	123
124	07/22/2021		N271	Response to Amendment under Rule 312			0
127	07/13/2021		N084	Issue Fee Payment Verified			0
123	07/13/2021		A.NA	Amendment after Notice of Allowance (Rule 312)			0
122	07/13/2021		N084	Issue Fee Payment Verified			0
121	07/13/2021		IFEE	Issue Fee Payment Received			0
120	06/04/2021		EML_NTR	Email Notification			0
118	06/04/2021		MM327-D	Mail PUB other miscellaneous communication to applicant			0
119	06/02/2021		CRFT	Sequence Forwarded to Pubs on Tape			0
117	06/02/2021		M327-D	PUB Other miscellaneous communication to applicant			0
115	05/26/2021		ELC_RVW	Electronic Review			0
114	05/26/2021		EML_NTF	Email Notification			0
113	05/26/2021	05/22/2021	MN/=.	Mail Notice of Allowance	3		86
116	05/25/2021		EIDC	Export to Initial Data Capture			0
112	05/24/2021		OAR	Office Action Review			0
111	05/24/2021		OAR	Office Action Review			0
110	05/24/2021		OAR	Office Action Review			0
109	05/24/2021		OAR	Office Action Review			0
108	05/24/2021		IREV	Issue Revision Completed			0
106	05/24/2021		N/=.	Notice of Allowance Data Verification Completed			0
105	05/24/2021		DOCK	Case Docketed to Examiner in GAU			0
103	05/18/2021		EX.A	Examiner's Amendment Communication			0
102	05/18/2021		EX.R	Reasons for Allowance			0
101	05/18/2021		CNTA	Allowability Notice			0
104	05/07/2021		EXET	Interview Summary - Examiner Initiated - Telephonic			0
100	05/07/2021		IDSC	Information Disclosure Statement considered			0
99	05/07/2021		IDSC	Information Disclosure Statement considered			0
107	05/04/2021		DVER	Document Verification			0
98	04/26/2021		EML_NTR	Email Notification			0
97	04/26/2021		MEXIN	Mail Examiner Interview Summary (PTOL - 413)			0

95	04/21/2021		EXIN	Interview Summary Record			0
96	04/19/2021		EXAT	Interview Summary - Applicant Initiated - Telephonic			0
94	03/30/2021		FWDX	Date Forwarded to Examiner			0
93	03/26/2021	01/22/2021	SA..	Supplemental Response		93	86
92	02/24/2021		EML_NTR	Email Notification			0
91	02/24/2021		MEXIN	Mail Examiner Interview Summary (PTOL - 413)			0
89	02/19/2021		EXIN	Interview Summary Record			0
90	02/18/2021		EXAT	Interview Summary - Applicant Initiated - Telephonic			0
87	01/27/2021		FWDX	Date Forwarded to Examiner			0
86	01/22/2021		A...	Response after Non-Final Action			0
85	01/22/2021		M844	Information Disclosure Statement (IDS) Filed			0
84	01/22/2021		M844	Information Disclosure Statement (IDS) Filed			0
83	01/22/2021		WIDS	Information Disclosure Statement (IDS) Filed			0
82	01/22/2021		WIDS	Information Disclosure Statement (IDS) Filed			0
81	10/22/2020		ELC_RVW	Electronic Review			0
80	10/22/2020		EML_NTF	Email Notification			0
79	10/22/2020	08/17/2020	MCTNF	Mail Non-Final Rejection	56		67
78	10/08/2020		OAR	Office Action Review			0
77	10/01/2020		NAFA125	Add FAOM 1.25 Count			0
76	10/01/2020		NSFA100	Subtract FAOM 1.00 Count			0
75	10/01/2020		CTNF	Non-Final Rejection			0
74	10/01/2020		IDSC	Information Disclosure Statement considered			0
73	10/01/2020		IDSC	Information Disclosure Statement considered			0
72	10/01/2020		IDSC	Information Disclosure Statement considered			0
69	04/19/2020		FWDX	Date Forwarded to Examiner			0
66	04/19/2020		ABN9	Disposal for a RCE / CPA / R129			0
70	04/17/2020		M844	Information Disclosure Statement (IDS) Filed			0
68	04/17/2020		AMSB	Amendment Submitted/Entered with Filing of CPA/RCE			0
67	04/17/2020	01/21/2020	RCEX	Request for Continued Examination (RCE)		57	57
65	04/17/2020		XT/G	Request for Extension of Time - Granted			0
64	04/17/2020		WIDS	Information Disclosure Statement (IDS) Filed			0
63	04/17/2020		BRCE	Workflow - Request for RCE - Begin			0
71	12/11/2019		M844	Information Disclosure Statement (IDS) Filed			0
61	12/11/2019		WIDS	Information Disclosure Statement (IDS) Filed			0
62	10/31/2019		M844	Information Disclosure Statement (IDS) Filed			0
60	10/31/2019		WIDS	Information Disclosure Statement (IDS) Filed			0
59	10/18/2019		ELC_RVW	Electronic Review			0
58	10/18/2019		EML_NTF	Email Notification			0
57	10/18/2019		MCTFR	Mail Final Rejection (PTOL - 326)			0
56	10/14/2019		CTFR	Final Rejection			0
55	10/14/2019		IDSC	Information Disclosure Statement considered			0
54	10/14/2019		IDSC	Information Disclosure Statement considered			0
51	08/09/2019		FWDX	Date Forwarded to Examiner			0
53	07/29/2019		C614	New or Additional Drawing Filed			0
52	07/29/2019		M844	Information Disclosure Statement (IDS) Filed			0
50	07/29/2019	07/05/2019	A...	Response after Non-Final Action		44	44
49	07/29/2019		XT/G	Request for Extension of Time - Granted			0
48	07/29/2019		WIDS	Information Disclosure Statement (IDS) Filed			0
47	07/29/2019		WIDS	Information Disclosure Statement (IDS) Filed			0
46	04/04/2019		ELC_RVW	Electronic Review			0
45	04/04/2019		EML_NTF	Email Notification			0
44	04/04/2019	03/20/2019	MCTNF	Mail Non-Final Rejection	38		38
43	03/31/2019		CTNF	Non-Final Rejection			0
42	03/31/2019		IDSC	Information Disclosure Statement considered			0
41	03/30/2019		IDSC	Information Disclosure Statement considered			0
39	01/06/2019		FWDX	Date Forwarded to Examiner			0
40	11/20/2018		M844	Information Disclosure Statement (IDS) Filed			0
38	11/20/2018		ELC.	Response to Election / Restriction Filed			0
37	11/20/2018		WIDS	Information Disclosure Statement (IDS) Filed			0
36	09/24/2018		ELC_RVW	Electronic Review			0
35	09/24/2018		EML_NTF	Email Notification			0
34	09/24/2018	03/30/2018	MCTRS	Mail Restriction Requirement	178		133.4
33	09/18/2018		CTRS	Restriction/Election Requirement			0
30	05/29/2018		CLTI	Close TI			0
29	05/29/2018		DOCK	Case Docketed to Examiner in GAU			0
28	05/28/2018		TI1050	Transfer Inquiry to GAU			0
27	02/01/2018		EML_NTR	Email Notification			0
26	02/01/2018		EML_NTR	Email Notification			0
20	02/01/2018		M903	Notice of DO/EO Acceptance Mailed			0
19	02/01/2018		FLRCPT.U	Filing Receipt - Updated			0
24	01/31/2018		COMP	Application Is Now Complete			0
23	01/31/2018		OIPE	Application Dispatched from OIPE			0
22	01/31/2018		FTFS	FITF set to YES - revise initial setting			0

15	08/10/2017		EML_NTR	Email Notification			0
14	08/03/2017		CCRDY	Application ready for PDX access by participating foreign offices			0
13	08/03/2017		PG-ISSUE	PG-Pub Issue Notification			0
25	06/13/2017		371COMP	371 Completion Date			0
21	06/13/2017		PTA.RFE	Patent Term Adjustment - Ready for Examination			0
18	06/13/2017		ADDFLFEE	Additional Application Filing Fees			0
16	06/13/2017		OATHDECL	A statement by one or more inventors satisfying the requirement under 35 USC 115, Oath of the Applic			0
12	04/24/2017		ELC_RVW	Electronic Review			0
11	04/24/2017		EML_NTF	Email Notification			0
10	04/24/2017		EML_NTR	Email Notification			0
9	04/24/2017		M923	371 Supplemental Fees Missing - Form M923			0
8	04/24/2017		FLRCPT.O	Filing Receipt			0
6	04/12/2017		WIDS	Information Disclosure Statement (IDS) Filed			0
5	04/11/2017		SMAL	Applicant Has Filed a Verified Statement of Small Entity Status in Compliance with 37 CFR 1.27			0
4	02/09/2017		CRFE	CRF Is Good Technically / Entered into Database			0
133.4	01/30/2017		COMMN	Commencement Date			0
0.5	07/29/2015		NEFILE	International Filing date			0

Export to: Excel

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO.	: 11,098,303 B2
APPLICATION NO.	: 15/329683
DATED	: August 24, 2021
INVENTOR(S)	: Zhuang et al.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

On the Title Page:

The first or sole Notice should read --

Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 319 days.

Signed and Sealed this
Twentieth Day of December, 2022

Katherine Kelly Vidal

Katherine Kelly Vidal
Director of the United States Patent and Trademark Office