PURIFICATION OF NANOREPORTER SCAFFOLD USING BIOTIN AND MAGNETIC BEADS WITH STREPTAVIDIN.

The fifth step occurs before attaching the flaps, where excess patch units that have not annealed to the M13mp18 strand are separated from the double stranded oligonucleotide scaffold. A purification tag with a 5 nucleotide base homologous region to some of the patch units' complementary 5 nucleotide base overhang is annealed to 'hook' the scaffold. Biotinylated oligonucleotides are annealed to the 'purification tag' and magnetic beads with streptavidin are used to capture the scaffold using the biotinylated oligonucleotides. Excess patch units are removed with the supernatant. The scaffold melts off of the magnetic beads into solution for recovery.

ANNEAL THE D-BIOTIN CATCHERS TO THE PURIFICATION TAGS

Anneal the D-Biotin catchers to the purification tags on the nanoreporter (making 2X to amount of D8-flap positions available in solution, which is 2X to M13, or 4X final): 0.5 pmol X 25 hook oligonucleotide positions (5 multiplied by 5), 4X makes 50 pmols translates to 0.50 µl of 100 pmol/µl D-biotin, add 0.5 µl (D, E, F) - Biotin (at 100 µM) to sample, mix and incubate at room temperature for 30 minutes.

PURIFICATION PROTOCOL TO WASH OFF UNATTACHED PATCH UNITS FROM DOUBLE STRANDED SCAFFOLD.

Anneal F-hook oligonucleotides in a 25 fold excess to nanoreporters in 5X SSC for 30 min at room temperature. Pipet 200 µl DynaBead MyOne Streptavidin™ bead solution into 1.5 ml tubes, place on magnet and remove supernatant. Wash twice with 5X SSC by resuspending and clearing with magnet as in step above. Add 80 µl of sample in 5X SSC (80 fmoles of sample in this example). Resuspend well, by placing on vortex for 15 minutes. Clear solution with magnet and transfer supernatant to fresh tubes for later gel analysis. While on magnet, wash pellets (do not resuspend) with 80 µl TE by pipeting over pellet three times with the same 80 µl volume originally added. Remove wash, place in freshly "washed" tubes for analysis. Heat up TE buffer to 45°C, add 80 µl to each pellet and resuspend. Place tubes on 45°C heat block for 15 minutes, pipetting up/down once to insure

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 984

beads remain suspended. Immediately clear product with magnet while warm and save. The majority of purified nanoreporters should be present in this product eluted at 45°C.

ANNEALING AND LIGATION OF FLAPS TO SCAFFOLD.

The sixth step involves split flap oligonucleotides which are annealed to the scaffold to make a 'covered scaffold.' Purification with magnetic beads is performed afterwards to remove excess split flaps. Ligation of the covered scaffold is done using T4 ligase to increase the stability of the structure. Only one type of flap is needed per fluorescent dye. Flaps are either 95 or 100 bases in length and have regions complementary to the patches, to labeled oligonucleotides and to each other. Each flap has 15 base repeating sequences for binding to labeled oligonucleotides. The repeat sequences are based on Lambda sequences that have been analyzed to remove any palindromes and hairpin structures.

Conditions for annealing the flaps are as follows. The sequence on the flaps that corresponds to the patch is 5 nucleotide base pairs long, and therefore the flaps anneal specifically to the patches even at high salt concentrations. The ratio of flaps to patches is 2:1. In order to increase stability at high temperatures, ligation of patches to each other and the flap to the patches may be carried out in the same reaction.

1) Quantify the purified scaffold using a spectrometer at A260 nm. Calculate the volume needed for appropriate amount of nanoreporter to prepare. For this example we used 110 ng or 0.023 pmol, reading at A260 nm shows 7.7 ng/µl, or 14.3 µl for 110 ng.
2) Setup ligation reaction as follows (volume will vary, depending on the purification and scale). Currently using 1.5X flaps to patches, calculate accordingly. For this example, there are four different fluorescent dyes (color) labeled A, B, C, and D and 8 different positions or regions where dye-labeled nucleic acids can bind on a nanoreporter. The number of positions for each color (in this case 1-4) multiply by 9 multiply by 1.5 moles of scaffold = moles of flaps to use.

For the nanoreporter with fluorescent dye in the sequence/positions [ABABCBAD]:

ABABCBAD =

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 985

A:40.5 x .023 = .93 pmol; vol: .93 µl of SF (split flap)-AL at 1 µM

.93 µl of SF-AR at 1 µM

B:40.5 x.023 = .93 pmol; vol: .93 µl of SF-BL at 1 µM

.93 µl of SF-BR at 1 µM

C:13.5 x .023 = .31 pmol; vol: .31 µl of SF-CL at 1 µM

.31 µl of SF-CR at 1 µM

D:13.5 x .023 = .31 pmol; vol: .31 µl of SF-DL at 1 µM

.31 µl of SF-DR at 1 µM

Ligation reaction (25 µl total) consists of : Split Flaps (see above; 4.96 µl, or ~5 µl total), 14.3 µl of MODB-Scaffold at 0.0016 pmol/µl, 2.5 µl 10X T4 ligation Buffer, 2.2 µl NanoPure H_2O and 1 µl T4 ligase. Incubate tubes 5 minutes at 45°C. Move to 37°C water bath, inc. for 5 minutes. Add 1 µl T4 ligase to samples. Incubate for additional 1 hour at 37°C. Freeze immediately, or heat at 75°C for 5 minutes to kill T4 ligase.

LIGATION OF TARGET-SPECIFIC SEQUENCES TO NANOREPORTERS

The seventh step involves ligation of a target-specific sequence to the nanoreporter. A DNA target-specific sequence is designed to be complementary to the target molecule, which can be RNA (*e.g.*, mRNA) or DNA (*e.g.*, cDNA or genomic DNA). The target-specific sequence can be from 35, 60 or 70 nucleotide bases in length. The target-specific sequence can be ligated to the scaffold using a single stranded overhanging region on the covered scaffold. The scaffold with a single type of target-specific sequence can be manufactured separately and then mixed to form libraries.

NANOREPORTER CONSTRUCTION

Addition of oligonucleotides to a nanoreporter can be done at any point during the construction of a nanoreporter. In certain aspects of the present invention, a labeled oligonucleotide is 15 nucleotide bases long. On the 5′ end, a single fluorophore dye is attached. Oligonucleotides with a particular fluorophore dye will generally have the same sequence. These labeled oligonucleotides bind to the repeat sequences of the split flaps. Fluorophores best suited for this example include but are not limited to Alexa 488, cy3,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 986

Alexa 594, and Alexa 647. The 15 nucleotide base length holds the fluorophores far enough apart so that they cannot quench each other and ensure that the labeled nucleic acids will be stable (will not melt off complementary strand) at conditions in the visualization process. Labeled oligonucleotides are stable at 40°C. This short length also allows for packing a large number of fluorescent dyes onto the flaps. In certain aspects of the invention, labeled oligonucleotides are introduced during the target sample processing.

ATTACHMENT OF NANOREPORTERS TO TARGET MOLECULES

Nanoreporters can be attached to target molecules using any means known to one of skill in the art. In an exemplary embodiment, dual nanoreporters are hybridized to target molecules by mixing 250 pmols each of both the first probe and the second probe with 125 pmols of target. The total volume is adjusted to 4 μl and a final concentration of buffer of 5X SSC. This mixture is incubated in a covered PCR tube overnight at 42 degrees to allow hybridization to occur.

SURFACE ATTACHMENT

Once the nanoreporters are attached to both target molecule and corresponding labeled nucleic acids, *i.e.*, nucleic acids attached to label monomers, they are attached to a surface and stretched to resolve the order of signals emitted by the label monomers and thus identify the target molecule. In this example, the nanoreporters are stretched to spatially resolve their fluorescent dye codes which correspond to a particular target molecule. The nanoreporters are stretched by attaching one end to a surface (in this example - a coverslip, see preparations below). Two methods for surface attachment may be used: A) streptavidin coated slides from Accelr8 Corporation with the nanoreporters being biotinylated and B) biotin coated slides with the nanoreporters having streptavidin. In buffer, the nanoreporters are brought into contact with the active surface and allowed to incubate for a period of time. The reaction is performed in flow cells which were made from PDMS molded in etched silicon wafers to make the channels. Metal tubing is used to core wells at the ends of the channels for buffer and sample insertion. Channel dimensions are 0.5 mm or 1 mm wide and 54 μm high. Once the sample has been loaded into the flow cell lane and incubated, the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 987

nanoreporters should be attached. Nanoreporters can be stretched either by applying a voltage or by removing the liquid with a receding meniscus leaving the strings stretched and dry.

PREPARATION OF SURFACE AND ASSEMBLY OF DEVICE

The binding surfaces (Accelr8 brand Streptavidin-OptiChem, coated coverslips) are shipped in units of 5 surfaces per slide container, and each container is enclosed with a package of silica dessicant in a foil pouch. The pouches are stored at -20°C until use.

To prepare the surface for binding, a pouch is first pulled from the freezer and allowed to come to room temperature over several minutes. If previously unopened, the pouch is then sliced along one edge to form a slit, and the container of surfaces is removed. Upon removal of the required surface, the container is replaced in the pouch with its dessicant, the slit is sealed closed with a strip of packaging tape, and the pouch is replaced in the freezer.

The surface is then lightly rinsed with a stream of Nanopure water (Barnstead Nanopure Diamond) and soaked for 10 minutes in 0.2 μm-filtered 1X PBS in a clean, slotted Coplin Jar. After soaking, the surface is dipped in Nanopure water and dried by blowing filtered nitrogen across the surface edge.

The PDMS device used to mate with the surface and provide localization of the sample is cleaned just before use by applying cellophane tape to the PDMS surface and then peeling away dust or other particles which may have become attached during storage. The binding side of the Accelr8 surface is laid face-up, and the clean PDMS structure is centered, channel side down, on the surface. PDMS adheres readily to coated glass, and no further attachment mechanism is necessary.

SAMPLE BINDING AND WASHING

The sample is bound to the surface by first applying a 5 μL drop of the sample (currently diluted in 100 mM sodium borate buffer, pH 9.8) in one well of the chosen lane. The drop should just touch the point at which the channel joins the well (some sample may

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 988

wick into the channel at this point). The channel is filled, and binding is equalized throughout the channel, by pulling the droplet through the channel to the opposite well using a very weak vacuum (< 2 kPa). The process is repeated for the other samples in their respective lanes. Excess fluid is then removed from the wells, the wells are taped to reduce evaporation, and the device is incubated at room temperature in the dark for 20 minutes.

After binding, the tape is removed, and the top well of each lane is filled with 100 μL of the borate buffer described above. About 20 μL of that buffer is pulled through the channels to the other wells using the vacuum, and the process is repeated once. All borate buffer is then removed from all wells, and the top well is filled with 1 X TAE, pH 8.3. About 50 μL TAE is pulled through the channel, then all TAE is removed and the well is refilled. The process is repeated three times, for a total of about 150 μL of TAE rinse. Finally, all wells are filled with 100 μL 1X TAE.

ELECTROSTRETCHING

The bottom of the coverslip/PDMS device is spotted with immersion oil and placed on the microscope. Electrodes are inserted into the wells on opposite ends of the first PDMS channel (negative electrode in top well, positive in bottom). The first image of the channel will be taken close to the bottom well; the microscope stage is adjusted so that the area of interest is in focus.

Voltage (200 V) is then applied across the channel. Voltage is supplied by a DC power supply (Agilent E3630A) and amplified 100X through a home-built amplifier. After the current is applied, focus is readjusted, and the imaging process begins.

The electrostretching and imaging process is then repeated with the remaining channels. Image the bindings.

LIGHT SOURCE FOR THE FLUORESCENT DYES ON THE NANOREPORTER

In using an arc lamp as a light source, the best fluorophore selection is the brightest types without leading to fluorescent overlap such as Alexa 488, Cy3, and Alexa 594. Weaker fluorescent dyes such as Alexa 647 and Cy5.5 may also be used.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 989

FILTERS TO IMAGE THE FLUORESCENT DYES ON THE NANOREPORTER

For the selected fluorophores Alexa 488, Cy3, Alexa 594 and Alexa 647 there may be an overlap between the Cy3 and Alexa 594. However, custom ordering an emission filter with a bandwidth of 572-600 nm minimizes the overlap.

MICROSCOPE AND OBJECTIVE LENS TO IMAGE THE NANOREPORTERS

The microscope model used was the Nikon Eclipse TE2000E from Nikon Corporation using the inverted fluorescence imaging station which has 6 filter cassettes that allow the selection of fluorescent emission from multiple fluorescent dye candidates. For the selected dyes, the optical resolution required is about 400 nm for all the wavelengths (500-700 nm). The selected objective lens is the Nikon Plan Apo TIRF lens which has a NA of 1.45 and magnification of 60. The optical resolution is ~210-300 nm for different wavelengths.

EXAMPLE 2: PATCH/FLAP NANOREPORTER MANUFACTURING PROTOCOL

This example demonstrates another way of making a nanoreporter which consists of a single stranded linear M13mp18 viral DNA, oligonucleotide patch units and long flaps.

Nanoreporter label units were successfully generated using methods substantially as described in this example.

Pre-phosphorylated patch units and flaps are added together with the M13mp18 DNA vector and ligated together. After the ligation of the flaps to the patch units which are ligated to the M13mp18 DNA, the BamH1 enzyme is introduced to linearize the vector.

Prepare a batch of nanoreporters starting with 5 μg of M13mp18 as a scaffold. The hybridization may be scaled up accordingly to the desired amount. This process will take about 1-2 days to complete.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 990

Materials:

Qty	Item	Vendor
20	250 ug/µl M13mp18 viral ssDNA	New England Biolabs
27 µl	0.74 pmol/µl Oligonucleotide Patch Unit Mix	IDT
8 µl	Long Flap Oligonucleotide A 100 pmol/µl	IDT
8 µl	Long Flap Oligonucleotide B 100 pmol/µl	IDT
0.5 µl	Flap patch Oligos at 100 pmol/µl from plates #529916 and #610591	IDT
31 µl	T4 Ligase 10x buffer	Fermentas
19 µl	T4 Ligase	Fermentas
15 µl	Optikinase 10x buffer	USB
4.2 µl	100 mM ATP	ANY
5 µl	Optikinase Enzyme 10 units/µl	USB
1 µl	BamH1 oligonucleotide 10 pmol/µl	IDT
20 µl	BamH1 10x buffer	Fermentas
3 µl	BamH1 Enzyme 10 units/µl	Fermentas

Preheat water bath to 37°C and 55°C before beginning protocol. Make sure buffers are all well mixed and thawed before using. A work plate should be available and labeled with the ordered oligos from IDT in plates #529916 and #610591. Take these two plates out and thaw at room temperature for 0.5-1 hours and spin down contents before removing the tape that covers the wells. Four separate reactions will be set up in 1.5 ml eppendorf tubes using specific oligonucleotides from these plates. To begin label these four separate tubes with roman numerals on their caps. Columns 5 and 6 A through H are for

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 991

reaction I, Columns 7 and 8 A through H are for reaction ii are all found in plate #529916. Columns 1 and 2 are for reaction iv, and Columns 3 and 4 are for reaction iii.

Flap Ligations (Step A): Label four separate 1.5 ml tubes with roman numerals i through iv (mentioned above). Add the reagents below accordingly to each 50 µl reaction containing: 5 µl 10x ligase buffer, 0.5 µl/oligonucleotide from designated wells from plates #529916 and #610591, 4 µl Long Flap Oligo/reaction (A or B) for reactions I, ii and iv. 3 µl of LF for area iii, 29 H_2O for reactions I,ii and iv. 32 µl H_2O for reaction iii, and 4 µl T4 ligase. Preanneal oligos in this mix without the ligase at 37°C for half an hour. Add ligase as last reagent and allow to ligate at room temperature for at least four hours. Product concentration is 1 pmol/flap/µl.

Flap Ligation Phosphorylation (Step B) Label four separate 1.5 ml tubes with roman numerals again, one through four with a P inside a circle to designate that the products are phosphorylated. Add the following reagents to the corresponding tube: 10 µl/Flap ligation reaction (take 10 µl/flap ligation reaction above), 2.5 µl Optikinase buffer, 0.5 µl 100 mM ATP, 11.5 µl H_2O, and 0.5 µl Optikinase enzyme. Incubate at 37°C for 1 hour. Product concentration 0.4 pmol/flap/µl.

Oligonucleotide Patch Unit Phosphorylation (Step C) 27 µl Oligonucleotide Patch Unit mix 0.74 pmol/µl, 5 µl 10x buffer, 1µl 100 mM ATP, 3 µl Optikinase enzyme, and 14 µl H_2O. Once reagents are all together gently mix the solution by flicking the tube a few times and spin down. Incubate at 37°C for 1 hour.

Hybridization to M13mp18 scaffold (Step D) In a new 1.5 ml tube add the following reagents: 20 µl M13mp18 at 250 ng/µl, 27 µl Phosphorylated Oligonucleotide Patch Units 0.4 pmol/µl (Step C), 12.5 µl/ Phosph. Flap Ligation (Step B) preheat at 55°C for 5 minutes and put on ice, 11 µl 10x ligase buffer and heat entire mixture at 55°C for 1 minute. Hybridize mixture at 37°C for at least 4 hours.

Ligation (Step E) Spin down eppendorf contents. Add 1.2 µl 100 mM ATP and 3 µl T4 ligase. Gently mix contents by flicking the tube, then spin down.

BamH1 Digest (Step F): 1 µl of 10 pmol BamH1 oligo, 20 µl 10x BamH1 buffer and hybride at 37 ~1 hour. Adjust volume to 200 µl. Add 3 µl BamH1 enzyme. Incubate at 37°C for 1 hour.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 992

First step: start by adding 20 µl of M13mp18 (NEB 250 µg/ml) to a clean 1.7 ml eppendorf tube. Take 5 µl of Phosphorylated Flap ligation reaction and preheat it at 70 for 2 minutes and immediately put on ice. Add the 5 µl of each Phosphorylated Flap Ligation reaction (1 pmol/flap/µl) to the tube and gently mix by pipetting a few times. Incubate the eppendorf tube at 37°C for 1 hour.

Second step: put 13.5 µl Oligonucleotide Patch Unit Mix (0.74 pmol/µl) and 1 µl of Acrydite Mix (10 pmol/µl) in a new eppendorf 1.7 ml eppendorf tube. Add 5 µl 10x Optikinase buffer, 1 µl 100 mM ATP and 27.5 µl H_2O. Mix gently by pipetting the solution. Add 2 µl Optikinase enzyme, gently mix by pipetting and incubate at 37°C for 1 hr.

Third step: take the phosporylated oligos rxn and add it entirely to the contents of the M13mp18+Flaps Hybridization. The reaction is mixed gently by pipetting and it is allowed to incubate at 30°C for 1 hour. After the hybridization is complete adjust the ATP by adding 1 µl (100 ATP) to the reaction.

Fourth step: spin down contents in eppendorf tube and add 4 µl T4 Ligase enzyme (5 units/µl), mix gently by pipetting. Incubate at room temperature for at least four hours. Add 1 µl BamH1 oligonucleotide (10 pmol/µl) to hybridize at room temperature while ligation is taking place.

Fifth step: digest ligation reaction by adding 4 µl BamH1 enzyme (5 units/µl), mix gently by pipetting and incubate at 37°C for 1 hour. Once the incubation period is over. Take an aliquot of 500 ng for QC.

Sixth step: treat with Psoralen, UV or DMPA light for 15 minutes.

Calculations include:

5 µg of M13 = 20 µl stock from New England Biolabs = 2 pmols

Oligonucleotide mix: 180-34 flap areas - 10 Acrydite modified Oligos = 0.74 pmol/oligo

10 pmols/oligonucleotide = 13.5 µl = 1350 pmols

Optikinase 1 unit converts 1 nmol of phosphate to ends - use excess. 4 µl of Optikinase was used.

SEQ ID NO: 1 = M13mp18.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 993

EXAMPLE 3: Protocol For Production Of RNA Nanoreporters

Nanoreporters were generated and successfully employed to detect target molecules using methods substantially as described in this example. An example of target detection using such this method is shown in Figure 6.

SCAFFOLD PRODUCTION

Single-stranded circular M13mp18 DNA (USB Corporation) is annealed to a 10-fold molar excess of an oligonucleotide complementary to the Bam HI recognition site (Bam Cutter oligo) and cut with Bam HI restriction enzyme to yield a linear single-stranded DNA backbone. An oligonucleotide complementary to the Bam Cutter oligonucleotide (anti-Bam oligonucleotide) is subsequently added in 50-fold excess to the Bam Cutter oligonucleotide to sequester free Bam Cutter oligonucleotide and thus prevent recircularization of the M13 during later steps.

The linear M13 molecule serves as a scaffold onto which RNA patches, or RNA segments, with incorporated fluorophores can be annealed.

PCR TO FORM DOUBLE-STRANDED POSITIONS ON THE M13 SCAFFOLD

Ten sets of oligonucleotide primer pairs were designed to create 10 different regions along the M13 scaffold. Each pair contains one primer which has a T7 RNA polymerase promoter at the 5′ end. Regions 2-7 are designed to be 900 bases (approximately 300 nm) long, as this is the approximate size of a diffraction-limited spot (the smallest spot that can be achieved with standard optics). Regions 1 and 8 have both long and short versions: the long versions cover the whole 900-base region, while the short versions cover only a portion of the 900-base region to allow a target-specific sequence to be ligated. Thus a target-specific sequence can be attached to either end. The ends can also be used for attachment of anchors or tags.

PCR is performed using Taq polymerase and 0.5 ng of double-stranded M13mp18 (USB Corporation) as a template. Reactions are cleaned up using a Qiaquick purification kit from Qiagen. Each PCR reaction yields a double-stranded fragment

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 994

corresponding to one specific segment as illustrated below. These fragments are used as templates for the *in vitro* transcription of the RNA segments.

IN VITRO TRANSCRIPTION TO PRODUCE DARK RNA SEGMENTS

Using the PCR products described above as double-stranded templates, RNA segments are generated using an *in vitro* transcription kit from Ambion (Megascript T7 kit). The products of the transcription reactions are purified (including treatment with DNAse I to remove template) using a RNeasy Kit from Qiagen.

IN VITRO TRANSCRIPTION TO PRODUCE RNA SEGMENTS MODIFIED WITH AMINOALLYL GROUPS

Using the PCR products described above as double-stranded templates, RNA segments for later dye-coupling are generated using an *in vitro* transcription kit from Ambion (MessageAmp aRNA kit). Aminoallyl-modified UTP nucleotides are incorporated into the RNA segments during transcription. The products of the transcription reactions are purified (including treatment with DNAse I to remove template) using a RNeasy Kit from Qiagen.

DYE COUPLING OF AMINOALLYL RNA SEGMENTS TO PRODUCE COLORED RNA SEGMENTS

20-100 μg of aminoallyl-modified RNA segment is coupled with NHS-ester dyes using Ambion Aminoallyl Labeling Kit. Dyes used include Alexa 488, Alexa 594 and Alexa 647 (Invitrogen/Molecular Probes) as well as Cy3 (Amersham).

Each segment is made separately in 4 colors so that each position on the scaffold can be filled with a segment in any of the four colors; thus different colors can be added at different positions to create many unique color combinations.

In this particular embodiment, adjacent segments must be of different colors or there may be dark segments interspersed so that each segment is detected as an individual 'spot'. Dark segments may be used as part of the nanoreporter code.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 995

ASSEMBLY OF THE LABEL MOLECULE

Segments for each position are annealed in a 2:1 ratio of segment to M13 scaffold in 1X SSPE buffer at 70°C for 2 hours.

An assembled nanoreporter with labeled RNA segments is depicted in Figure 3A-3B. Figure 3A depicts a nanoreporter in which only alternate "spots" (1, 3, 5 and 7) are labeled, and Figure 3B depicts a nanoreporter in which every spot is labeled.

EXAMPLE 4: DETECTION OF TARGET (S2) RNA AND DNA MOLECULES USING AN RNA NANOREPORTER/GHOST PROBE COMBINATION

SYNTHESIS OF PROBE AND TARGET OLIGONUCLEOTIDES

S2 DNA target oligonucleotide was synthesized and purified by polyacrylamide gel electrophoresis (Integrated DNA Technologies). S2 RNA target molecules were generated by *in vitro* transcription of PCR products corresponding to region of cloned SARS coronavirus gene (Invitrogen) using an Ambion Megascript™ kit per manufacturer's instructions. The S2 ghost probe (Figure 6A (i)) was complementary to a specific 50-base region of the S2 target sequence (S2-*a*) and was synthesized with a biotin-TEG monomer at the 5′ end and purified by high performance liquid chromatograpy (Integrated DNA Technologies). A second oligonucleotide with 50 bps complementary to the S2 target (S2-*b*) plus 9 bp of an additional sequence used for ligation to the M13 scaffold (59bp total) was synthesized and purified by HPLC (Integrated DNA Technologies). Note that S2-*a* and S2-*b* target regions were not overlapping.

NANOREPORTER SYNTHESIS

Oligonucleotide S2-*b* was ligated to the 5′ end of linearized M13 [Figure 6A (iii)], and the resulting product was purified away from residual unligated oligonucleotide by size-exclusion filtration through a YM100 filter (Millipore) per manufacturer's instructions. Amino-allyl-modified RNA segments complementary to M13 is positions 2, 4, 6, and 8 (SEQ ID NOs:) (Figure 1C) were generated from *in vitro*-transcription of DNA templates (PCR products) via the Ambion Megascript™ kit per manufacturer's instructions. The segments

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 996

were then coupled to NHS-ester-modified Alexa 647 dye (Invitrogen) per Ambion's instructions (amino allyl MessageAmp™ II aRNA kit). RNA segments corresponding to positions 1, 3, 5, and 7 of the M13 scaffold (Figure 1C) were generated as unmodified *in vitro*-transcribed RNAs from DNA templates as described above. Assembly of the nanoreporter was carried out by annealing 10 fmol/µl of each of the eight segments to 5 fmol/µl of the M13-S1-*b* scaffold for 2 hours at 70°C in 1X SSPE buffer (150 mM sodium chloride, 10 mM sodium phosphate, 1 mM EDTA). The final product was a nanoreporter with 4 segments labeled with A647 (red) interspersed with dark segments.

HYBRIDIZATION CONDITIONS

Hybridization of nanoreporters and ghost probes to target were carried out under the following conditions: 5X SSPE (750 mM sodium chloride, 50 mM sodium phosphate, 5 mM disodium EDTA), 40 pM ghost probe (attachment oligonucleotide S2-*a*), 40 pM Nanoreporter S2-*b*, 100 ng/µl sheared salmon sperm DNA, 5X Denhardt's solution and 0.1% Tween. Final target concentrations were 20 pM S2 DNA target (Figure 6B) and 1pM S2 RNA target (Figure 6C). No target was added to the negative control (Figure 6D). The hybridization reaction was incubated at 65°C for at least 16h.

Hybridization reactions were diluted 1:2 with 100 mM Borate buffer solution (pH 9.8) and introduced into a flow cell channel and bound to a streptavidin-coated coverslip forming the bottom of the channel (Streptavidin-OptiChem coverslips from Accelr8). Attachment to the slide by one end of the nanoreporter/target/ghost probe complex was achieved via interaction of the biotinylated ghost probe with the streptavidin surface. After rinsing the channel with additional borate buffer to remove excess reporters not bound to the surface, the buffer was exchanged with 1X TAE (40 mM Tris-acetate, 1 mM EDTA) and a current of 200V was applied to stretch out the nanoreporter/target complexes during image capture.

Images were obtained using a Leica DMI 6000B microscope with a 63X oil immersion objective (1.4 NA), Xcite-120 light source (Exfo), customized filter sets (Chroma Technologies), an Orca-ER CCD camera (Hamamatsu) and Metamorph data acquisition software (Molecular Devices).

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 997

As predicted, when the correct target molecule S2 hybridizes [Figure 6A (ii)] to both ghost probe [Figure 6A (i), S2-*a*] and S2-*b* target-specific nanoreporter [Figure 6A (iii)], the ghost probe/target/nanoreporter complex forms a single species that attaches to the slide and was visualized as 4 spots when exposed to 647 nm wavelength light (Figure 6B, 6C, and 6E). The amount of binding was dependent on the target concentration. There was no significant binding in absence of S2 target sequence (Figure 6D).

EXAMPLE 5: NANOREPORTER COMPRISING A MONOVALENT OR BIVALENT ANTIBODY FRAGMENT

Where a target molecule is a protein or polypeptide, a nanoreporter can be generated in which the nanoreporter scaffold is a nucleic acid and the target-specific sequence is a monovalent or bivalent antibody fragment.

Using routine methods, an antibody that recognizes a target molecule of interest is optionally digested with pepsin to generate F(ab´)2 fragments. The two parts of the antibody or the two F(ab´)2 fragments generated by the pepsin digestion are separated by mild reduction, for example with 2-mercaptoethylamine. This reduction separates either the antibody or the two F(ab´)2 fragments into two monovalent fragments with two sulfhydryl groups that can be functionalized.

A heterobifunctional crosslinking reagent (*e.g.*, m-Maleimidobenzoyl-N-hydroxysuccinimide ester from Pierce Biotechnology Inc.) is used to attach a maleimide to an oligonucleotide with an amine modification (which can be ordered from many sources, such as Integrated DNA Technologies). The NHS on the cross-linking reagent is reacted with the amine on the oligonucleotides to produce a maleimide-conjugated oligonucleotide.

This maleimide conjugated oligonucleotide is then reacted with one of the sulfhydryl groups on the antibody fragment. Due to steric limitations, it is preferable that only one oligonucleotide be attached to each fragment.

This monovalent or bivalent antibody fragment attached to an oligonucleotide can then be hybridized to a complementary sequence on a nanoreporter scaffold, to generate a reporter probe in which the target-specific sequence is an antibody sequence. Such a reporter probe can be used alone to detect the target molecule, or in conjunction with a ghost probe or

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 998

another reporter probe whose target-specific sequence is a monovalent or bivalent antibody or antibody fragment that binds to a different portion of the same target molecule.

EXAMPLE 6: HYBRIDIZATION OF 25 CELLULAR GENES TO 100 NG OF PLACENTAL TOTAL RNA USING NANOSTRING REPORTER SYSTEM

Detection and quantitation of 25 endogenous cellular genes was carried out in a single multiplexed hybridization reaction. In addition, three non-human control sequences were spiked into each reaction that corresponded to approximately 10, 100 and 300 copies per cell, respectively. A negative control hybridization was also performed in the absence of cellular RNA.

HYBRIDIZATION REACTION

Each sample was hybridized in triplicate. Final concentrations of the hybridization reagents were as follows: 1.12nM total Nanoreporters (28 individual Nanoreporters at 40 pM each), 1.12nM total ghost probe (28 individual ghost probes), 5X SSPE (pH 7.5), 5X Denhardt's reagent, 100 ng/μl sheared salmon sperm DNA, 0.1% Tween 20, 150 fM S3 spike DNA, 50 fM S4 spike, and 5 fM S6 spike. The final concentration of total placental RNA was 33 ng/μl. No total placental RNA was added to the negative control hybridizations. The final volume of the reaction was 30 μl. Reagents were mixed and incubated at 65°C in thermocycler block with heated lid for 20 hours.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 999

Master Mix	(1 Reaction)	(6 Reactions)
1.8X hybridization mix*	16.7 µl	100 µl
25 endogenous gene reporters (0.6 nM each)	2 µl	12 µl
25 endogenous gene ghost probes (0.6 nM each)	2 µl	12 µl
Control reporters (0.6 nM each)	2 µl	12 µl
Control ghost probes (0.6 nM each)	2 µl	12 µl
10X control target mix	3 µl	18 µl
H_2O	1.3 µl	8 µl
Total	29 µl	174 µl

*Hybridization mix (9X SSPE, 9X Denhardt's reagent, 180 ng salmon sperm DNA, 0.18% Tween 20)

Reactions	1	2	3	4	5	6
Master mix	29 µl	29 µl	29 µl	29 µl	29 µl	29 µl
100 ng/µl placental RNA	1 µl	1 µl	1 µl	0 µl	0 µl	0 µl
H_2O	0 µl	0 µl	0 ul	1 µl	1 µl	1 µl
Total Rxn volume	30 µl	30 µl	30 µl	30 µl	30 µl	30 µl

Incubate reactions in thermocycler with heated lid overnight (18 hours).

POST-HYBRIDIZATION PURIFICATION

Hybridization reactions were purified to remove unhybridized reporters using an oligonucleotide complimentary to ghost probe attached to magnetic beads (F-bead). Hybridization reactions were diluted 5 fold in 0.1% Tween 20 to bring the final salt concentration to 1X SSPE and the solution added to 30 µl of F-beads (prewashed 2 times in 150 µl of 1X SSPE/0.1% Tween 20). Hybridized complexes were allowed to bind to the beads at room temperature for 15 minutes with continuous rotation, washed once in 150 µl of 0.5X SSPE, and eluted in 25 µl of 0.1X SSPE for 15 minutes at 45°C.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1000

BINDING, STRETCHING, AND IMMOBILIZATION

The samples was prepared for binding by addition of 1 µl of 1/1000 dilution of 0.1 uM Tetraspec™ fluorescent microspheres (product # T7279, Molecular Probes) and 3 µl of 1M bis-tris propane (pH 9.0). Samples were loaded into a Nanostring fluidic device for attachment to Accelr8 Optichem® slide coated with streptavidin (product #TB0200). After loading, slide surface was washed once with 1X TAE and prepared for electrostretching by addition of 40 µl of TAE to each well. Attached complexes were stretched by applying 200V across the fluidic channel. After 1 minute the samples were immobilized in the stretched position by adding 60 µl of 500 mM of G-hook oligo solution to the well containing the negatively charged electrode while continuing to apply voltage for 5 minutes. After immobilization the TAE solution is removed and replaced with anti-photobleaching reagent for imaging.

IMAGING

Slides were imaged on Nikon Eclipse TE2000E equipped with a metal halide light source (X-cite 120, Exfo Corporation) and a 60X oil immersion lens (1.4 NA Plan Apo VC, Nikon). For each field of view, 4 images at different excitation wavelengths (480, 545, 580 and 622) were acquired with an Orca Ag CCD camera (Hamamatsu) under control of Metamorph software (Universal Imaging Corporation). Images were processed with custom image processing software.

DATA ANALYSIS

Raw data were extracted from processed images using custom software. Data were normalized to the average counts for control spikes in each sample. To determine if a gene was "detected" by the system, the counts obtained for each gene from hybridizations containing RNA were compared to counts obtained in hybridizations without RNA using a Student's t-test. Genes with p values <0.05 were determined to be detected. After background subtraction, the concentrations of cellular mRNA were estimated from the linear regression of the spike controls. These concentrations were converted to copies per cell using

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1001

the following assumptions: 1 cell contains 10 pg total RNA; each cell contains 300,000 mRNA molecules; final volume of the reaction is 30 µl.

RESULTS AND CONCLUSION

Table 3 below shows the results of the data analysis described above. These results show that using the nanoreporter technology described herein, it was possible to detect transcripts, such as CASP3, that are present at a concentration of less than 1 transcript/cell. Thus, the nanoreporter technology provides an exquisitely sensitive means of detecting and quantifying gene expression.

Table 3: Transcript Concentration and Abundances

Gene	Avg counts*	error (counts)	concen-tration (fM)	error (conc.)	calculated copies/cell	error (copy/cell)	Detected/ Not De-tected
							($p<0.05$)
GM2A	149	17	3.39	0.39	6.12	0.07	D
ATF4	68	2	1.55	0.06	2.80	0.01	D
CTNNB1	792	50	17.95	1.19	32.44	0.22	D
IRF1	221	20	5.01	0.47	9.05	0.09	D
STAT5A	120	11	2.72	0.25	4.91	0.05	D
CREG1	409	17	9.28	0.44	16.76	0.08	D
CASP3	13	1	0.30	0.03	0.54	0.00	D
CCL20	2	1	0.04	0.03	0.07	0.01	ND
NMI	115	2	2.61	0.07	4.72	0.01	D
XBP1	719	46	16.30	1.10	29.45	0.20	D
PCGF4	75	18	1.70	0.40	3.08	0.07	D
IFI27	747	41	16.94	1.00	30.61	0.18	D
TAF7	185	11	4.19	0.26	7.57	0.05	D
OAS3	74	9	1.68	0.20	3.03	0.04	D

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1002

Gene	Avg counts*	error (counts)	concen-tration (fM)	error (conc.)	calculated copies/cell	error (copy/cell)	Detected/ Not De-tected
							($p<0.05$)
C2	850	49	19.28	1.19	34.83	0.21	D
IL6	8	3	0.19	0.07	0.34	0.01	D
MyD88	94	6	2.13	0.14	3.85	0.03	D
HIF1A	130	7	2.95	0.17	5.33	0.03	D
APOA2	-1	2	-0.01	-0.05	-0.03	-0.01	ND
KISS	6825	130	154.79	4.52	279.65	0.82	D
ELK3	55	4	1.25	0.09	2.27	0.02	D
CBF2	72	3	1.64	0.07	2.96	0.01	D
IFI30	625	47	14.16	1.10	25.59	0.20	D
RELB	35	5	0.78	0.11	1.42	0.02	D
CTCF	103	3	2.35	0.09	4.24	0.02	D

*Normalized and background subtracted.

The hybridization methods described herein have been performed in single multiplexed reactions containing up to 120 different reporters with similar hybridization efficiencies and results.

EXAMPLE 7: CONSIDERATIONS REGARDING NANOREPORTER HYBRIDIZATION KINETICS

BACKGROUND

Solution hybridizations with a large excess of probe over target follow pseudo-first order kinetics. In this regime the speed of the reaction depends only on the probe concentration and not on the target concentration. For a two-probe, one-target strategy to provide accurate information on the concentration of a target in solution, the probes should both be present in excess of the target. The possible concentration range is preferably therefore bounded on the lower end by the concentration of the target. However, the useful

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1003

concentration range for the nanoreporter technology described herein is practically bounded on the lower end by the amount of time needed to perform the hybridization.

HYBRIDIZATION KINETICS

In preferred embodiments, target detection and quantification assays are performed in which the target (T) must hybridize to both a reporter probe (R) and a ghost probe (G) to be detected (for example by affinity selection and detection of complexes comprising only (R) and (G), which in turn only form complexes in the presence of (T)). Assuming that these reactions are irreversible, there are four possible elementary reactions that occur.

$$R + T \xrightarrow{k_1} RT$$

$$T + G \xrightarrow{k_2} TG$$

$$RT + G \xrightarrow{k_3} RTG$$

$$R + TG \xrightarrow{k_4} RTG\,.$$

Because RT and TG are intermediate complexes of two out of the three species, these four reactions can be simplified to

$$R + T + G \rightarrow RTG\,.$$

However, to quantitatively calculate the rate of production of RTG (the reporter-target-ghost probe complex), all four reactions must be considered. The differential equations describing the system are:

$$\frac{dC_G}{dt} = -k_2 C_G C_T - k_3 C_G C_{RT} \qquad \frac{dC_G}{dt} = -k_1 C_R C_T - k_4 C_R C_{TG} \qquad \frac{dC_T}{dt} = -k_2 C_G C_T - k_2 C_R C_T$$

$$\frac{dC_{TG}}{dt} = k_2 C_G C_T - k_4 C_R C_{TG} \qquad \frac{dC_{RT}}{dt} = k_1 C_R C_T - k_3 C_G C_{RT}$$

$$\frac{dC_{RTG}}{dt} = k_4 C_R C_{TG} + k_3 C_G C_{RT}$$

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1004

where C_R, C_T, C_G, C_{RT}, C_{TG}, and C_{RTG} are the concentrations of the various species, and k_1-k_4 are the kinetic constants for the four elementary reactions. Values for these kinetic constants when the probes and targets are complementary single-stranded molecules (*i.e.*, when there is no purification tag on the ghost probe and no reporter) can be calculated from data available in the literature (Wetmur, *J. Annu. Rev. Biophys. Bioeng.* 1976.5:337-361).

$$k = k_N \frac{\sqrt{L}}{N} \frac{\alpha_{salt}}{\alpha_{ref}}$$

In the above equation, k_N is the nucleation rate constant, L is the nucleic acid length (in base pairs), N is the nucleic acid complexity (equal to L for non-repetitive sequences) and a_{salt} and a_{ref} are corrections for salt concentration (Britten *et al.*, 1974, *Methods in Enzymology* 29E:363-406). In the nanoreporter systems described herein, the kinetic constants will depend on the sizes of the attached ghost probe tags and reporter probe. Without being bound by any theory, it is the inventors' belief that the kinetic constants will have the same dependence on length that an elementary reaction has on the diffusion constants of the reactants.

$$k = k_N \frac{\sqrt{L}}{N} \frac{\alpha_{salt}}{\alpha_{ref}} \frac{D_1 + D_2}{2D_{50}}$$

In the above equation D_1 and D_2 are the diffusion constants of the two reacting species (see the reactions above) and D_{50} is the diffusion constant of a 50-mer single-stranded DNA molecule. Assuming a 100-base single-stranded target, 100-base single-stranded ghost probe, and 7200-base double stranded reporter, the relevant kinetic constants are

$k_1 = 2.64 \times 10^5$ L/mol/s

$k_2 = 6.55 \times 10^5$ L/mol/s

$k_3 = 3.99 \times 10^5$ L/mol/s

$k_4 = 1.91 \times 10^5$ L/mol/s

Numerically solving the system of differential equations with these kinetic constants (assuming at least a 10-fold excess of probes over target) yields the prediction that 5 pM reporter and 5 pM ghost probe will drive hybridization to 10% of completion in an overnight reaction (16-18 hours). At concentrations lower than 5 pM, the amount of

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1005

completely hybridized molecules is likely impractical to measure. Thus, in a preferred embodiment, the lower concentration of a nanoreporter component (ghost probe and/or reporter probe) is 5 pM.

ENTANGLEMENT OF REPORTERS

As probe concentrations increase, theory predicts that hybridization kinetics speed up without bound - the only limit being the solubility of the probes. However, the reporter probe can be very large compared to the target-specific sequence in the nanoreporter systems of the invention. Without being bound by any theory, the inventors believe that by its attachment to the reporter probe the kinetics of the target-specific sequence are altered from classical solution hybridization kinetics. Because the reporter probe is a large, polymeric molecule, it can have long-lived interactions (entanglements) with other nanoreporters when they come into contact. At low concentration the probability of two polymers becoming entangled is small, but as the concentration and/or size of a polymer in solution increases, these interactions become more and more common. In the extreme case of very long molecules at very high concentration the polymers form a permanent network, or gel, in solution. For solution hybridization to occur, a probe (*e.g.*, a nanoreporter probe)/target pair must diffuse through solution until they contact one another and a hybridization nucleus forms. Classically, hybridization reactions are not diffusion limited because the translational diffusion of the molecules is faster than the nucleation of the hybridization (*i.e.*, the probe and target diffuse together and interact many times before a nucleation occurs). In dilute solution its large size will slow the translational diffusion of the reporter probe, but may not significantly affect the kinetics. At some intermediate concentration, the reporter probes take up almost all of the space in the solution, effectively forming a permanently entangled gel, and can no longer diffuse in solution. However, the ghost probe and the targets are smaller molecules that are believed to still diffuse through the entangled reporter probes, allowing hybridization to take place (although possibly at a slower rate). The inventors also believe that at some higher concentration the reporter probe in solution will also hinder the movement of the ghost probe and the targets to the point that the reaction becomes diffusion limited. This concentration (which is not quantitatively known

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1006

and depends upon the reporter probe structure, the ghost probe structure, and the target size) is the upper limit of the useful concentration range in the nanoreporter system, and can be empirically determined by one of skill in the art guided by the principles described herein.

LENGTH DEPENDENCE OF KINETICS

Since the limiting upper concentration for hybridization depends upon both the reporter structure and ghost probe structure (of which there are many possible variations), a theoretical framework to predict the permutations of useful concentration ranges is useful in the practice of the invention. Classical theory predicts that hybridization kinetics depend only on the size of the smaller probe. Theory would therefore predict that the size of the reporter will not play a role in the hybridization kinetics as long as both the target molecule and the ghost probe are significantly smaller. Theory then predicts that the rate of hybridization (for a constant target length) depends on $1/L^{1/2}$, where L is the length of the ghost probe, due to steric inhibition of hybridization. Consequently, the kinetics of hybridization will be faster with smaller ghost probes. As the ghost probe length increases, the hybridization rate should decrease as $1/L^{1/2}$. If a constant ghost probe length is assumed, then the range of reporter lengths and concentrations that will result in a measurable mount of hybridization events can be defined. Once a reporter size has been defined, then the approximate range of ghost probe sizes can be determined. This is an iterative process, but may give good starting points from which to gather data to generate detailed empirical guidelines, given that the theories that the inventors' rationale is based upon were generated from hybridization data in systems that do not employ a reporter probe.

ENTANGLEMENT THRESHOLD

A reporter probe is essentially a polymer in free solution, which behaves as a random coil. The volume occupied by a single reporter, V_p, can be calculated from polymer physics theories according to the Freely-Jointed Chain model (FJC, for a flexible polymer, such as single-stranded DNA or RNA) or the Worm-Like Chain model (WLC, for a stiff polymer such as double-stranded DNA or a reporter). For either model

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1007

$$V_p = \frac{4}{3}\pi R_g^3$$

where R_g is the radius of gyration. For the FJC

$$R_g = b\left(\frac{N}{6}\right)^{0.6}$$

where b is the segment length and N is the number of segments in the chain. For the WLC

$$R_g = \sqrt{\frac{1}{6}Nb^2 - \frac{b^2}{4} + \frac{b^2}{4N}\left(1 + \frac{1}{2N}\left(e^{-2N} - 1\right)\right)}$$

The entanglement threshold concentration is defined as the concentration where the entire volume of the solution is occupied by the reporters.

$$C^* = \frac{3}{4\pi R_g^3 N_A}$$

where N_A is Avogadro's number. Above this concentration it is assumed that the translational diffusion of the reporters is severely restricted. The entanglement threshold concentration varies with the reporter structure. As the reporter length increases, the entanglement threshold decreases (as $1/L^{1.5}$). From the equations above, the theoretical entanglement threshold for reporter probes with different spot sizes and different lengths can be calculated. The result of such calculations is shown in Figure 17, which shows that for a 7200 bp RNA/DNA hybrid reporter probe with 8 label attachment regions of about 900 bp each, the entanglement threshold is about 70 nM.

If both the target and the ghost probe are much smaller than the reporters, then they will most likely be free to diffuse through the solution even at these high concentrations of reporters. Initial data indicates that hybridization kinetics do not slow appreciably up to a concentration of 80 nM with a 7200-bp reporter probe, a 100-base target, and a 100-base ghost probe.

EFFECT OF ENTANGLEMENT THRESHOLD ON MULTIPLEXING

Assuming that the maximum concentration for reporters in a hybridization reaction is C^*, then the concentration of each reporter (specific to a particular target) is equal

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1008

to C^*/M, where M is the multiplex of the reaction (number of different targets being addressed simultaneously). Conversely, the possible multiplex level for a particular reporter structure can be calculated from the lower limit of probe concentration (C_p from kinetics ~ 10 nM) and the entanglement threshold

$$M = \frac{C^*}{C_p}$$

If the number of nanoreporter codes available does not depend on reporter probe size, then the multiplexing of the nanoreporter depends primarily on the reporter probe size and concentration (since it is much larger than the ghost probe). Because the ghost probe makes an insignificant contribution to entanglement during hybridization, it is the inventors' belief that the concentration of the ghost probe can be increased far above the concentration of the reporter probe. In Table 4 below, the maximum total ghost probe concentration ([G]) is set to 1000 nM for all reporter concentrations. This difference in concentration of ghost probe and reporter probe is an adjustable parameter. Preliminary experiments show that in a multiplex hybridization reaction with a 7200 bp reporter and 100b ghost, 40pM of each reporter probe and 200 pM of each ghost probe results in near complete hybridization in an overnight reaction.

OPTIMAL SIZE AND CONCENTRATION RANGES

Below in Table 4 is a summary of the optimal useful size and concentration ranges of the ghost probe and reporter probe at different multiplexing as approximated by the above theories. It is the inventors' belief that ghost probes up to about 200 bases will be practical for most applications

Table 4: Optimal size and concentration ranges of reporter probe, ghost probe and target, as well as multiplicity of probes, in the nanoreporter systems of the invention.

Reporter Length (bp)	Ghost Length (b)	Minimum [R] (pM)	Minimum [G] (pM)	Maximum [R](nM)	Maximum [G](nM)	Max Multiplex
2000	100	5	5	603	1000	114417

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1009

Reporter Length (bp)	Ghost Length (b)	Minimum [R] (pM)	Minimum [G] (pM)	Maximum [R](nM)	Maximum [G](nM)	Max Multiplex
2000	50	4	4	603	1000	161811
2000	200	7	7	603	1000	80905
3000	100	6	6	292	1000	45182
3000	50	5	5	292	1000	63897
3000	200	9	9	292	1000	31948
4000	100	7	7	178	1000	23912
4000	50	5	5	178	1000	33817
4000	200	11	11	178	1000	16908
5000	100	8	8	123	1000	14746
5000	50	6	6	123	1000	20854
5000	200	12	12	123	1000	10427
6000	100	9	9	91	1000	9988
6000	50	6	6	91	1000	14125
6000	200	13	13	91	1000	7062
7200	100	10	10	68	1000	6792
7200	50	7	10	68	1000	6792
7200	200	14	10	68	1000	6792
8000	100	11	11	57	1000	5444
8000	50	7	7	57	1000	7699
8000	200	15	15	57	1000	3850
10000	100	12	12	40	1000	3419
10000	50	8	8	40	1000	4835
10000	200	17	17	40	1000	2417

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1010

EXAMPLE 8: EXEMPLARY EMBODIMENTS FOR DUAL NANOREPORTER ASSEMBLY

This section describes an embodiment for assembly of a dual nanoreporter in which one probe is a ghost probe and the other probe is a reporter probe comprising color RNA segments assembled on an M13 backbone. The ghost probe is attached to a biotinylated F-hook and the reporter probe is attached to a biotinylated G-hook. The dual nanoreporter is hybridized to a biomolecular sample to detect and quantify a target molecule. The steps below do not have to be performed in the order presented. Moreover, each particular step represents a specific embodiment that may be combined with embodiments other than those presented below.

PREPARATION OF THE M13 SCAFFOLD

Single-stranded circular M13mp18 DNA (USB Corporation) is annealed to a 5-fold molar excess of an oligonucleotide complementary to the Bam H1 recognition site (Bam Cutter oligo) and cut with Bam H1 restriction enzyme to yield a linear single-stranded DNA backbone. An oligonucleotide complementary to the Bam Cutter oligonucleotide (anti-Bam oligonucleotide) is subsequently added in 50-fold excess to sequester free Bam Cutter oligonucleotide and thus prevent recircularization of the M13 during later steps.

The linear M13 molecule serves as a scaffold onto which RNA patches, or RNA segments, with incorporated fluorophores can be annealed.

ATTACHMENT OF A TARGET-SPECIFIC SEQUENCE TO THE SCAFFOLD

An oligonucleotide comprising a sequence (of, *e.g.*, 30-70 nucleotides) complementary to the target nucleic acid of interest, plus 9 bp of additional sequence used for ligation to the M13 scaffold, is generated and ligated to the 3′ end of the linearized M13 scaffold.

ATTACHMENT OF G-TAGS TO THE SCAFFOLD

A G-tag (*e.g.*, an oligonucleotide having the sequence 5′-AACATCACACAGACC AACATCACACAGACC AACATCACACAGACC

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1011

AACATCACACAGACC AGCCCTTTG-3′, which includes 4 copies of the complement of the G-hook 5′-GGTCTGTGTGATGTT-3′, followed by 9 bases of ligator sequence, and which is complementary to the G-hook) is attached to the 5′ end of the linearized single-stranded M13 backbone to allow for (1) purification of the reporter following ligation and/or annealing of segments; and (2) immobilization of the reporter once it is "stretched" on a solid surface. The sequence of the ligator for attaching G-tag to the 5′ end of single-stranded M13 which has been linearized at the BamH1 site can be 5′-CTCTAGAGGATCCAAAGGGCT-3′. The ligation reaction can be performed according to the following protocol to produce approximately 80 pmol of G-tag/M13 ligation product:

Materials:

[100 μM] anti-G4 tag oligo
[100 μM] anti-G4 tag ligator oligo
[80 nM] Linear single-stranded M13
[10X T4 DNA Ligase Buffer (Fermentas)
T4 DNA Ligase (Fermentas)
20X SSC (Ambion)
DEPC H_2O (Ambion)

Method:

1. Pre-anneal the G-tag and ligator:
   25uM 2:1 G/Glig in 1X SSC
   20 μl [100uM] G-tag Ligator
   40 μl [100uM] G-tag
   4 μl 20X SSC
   16 μl DEPC H_2O

* Anneal on the MJ Thermocycler

95°C, 3min; 72°C, 30sec, -1°C/cycle, x 68 cycles; hold at 4°C

2. Ligate the G-tag to the linear M13:
   64nM M13-G4 in 1X Lig Buffer

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1012

1000 µl [80 nM] Linear M13

80 µl [25uM] 2:1 G/Glig in 1X SSC

124 µl 10X T4 DNA Ligase Buffer

40 µl T4 DNA Ligase

* Ligate in an aluminum heat block covered with foil at 37°C for 2hr then at 65°C for 15 minutes to inactivate the enzyme.

PREPARATION OF RNA SEGMENTS

Ten sets of oligonucleotide primer pairs are designed to create 10 different regions along the M13 scaffold. Each pair contains one primer which has a T7 RNA polymerase promoter at the 5′ end. Regions 2-7 are designed to be 900 bases (approximately 300 nm) long, as this is the approximate size of a diffraction-limited spot (the smallest spot that can be achieved with standard optics). Regions 1 and 8 have both long and short versions: the long versions cover the whole 900-base region, while the short versions cover only a portion of the 900-base region to allow a target-specific sequence to be ligated. Thus a target-specific sequence can be attached to either end. The ends can also be used for attachment of anchors or tags.

PCR is performed using Taq polymerase and 0.5 ng of double-stranded M13mp18 (USB Corporation) as a template. Reactions are cleaned up using a Qiaquick purification kit from Qiagen. Each PCR reaction yields a double-stranded fragment corresponding to one specific segment as illustrated below. These fragments are used as templates for the *in vitro* transcription of the RNA segments.

Using the PCR products described above as double-stranded templates, RNA segments are generated using an *in vitro* transcription kit from Ambion (Megascript T7 kit). The products of the transcription reactions are purified (including treatment with DNAse I to remove template) using a RNeasy Kit from Qiagen.

LABELING OF THE RNA SEGMENTS

Using the PCR products described above as double-stranded templates, RNA segments for later dye-coupling are generated using an *in vitro* transcription kit from Ambion

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1013

(MessageAmp aRNA kit). Aminoallyl-modified UTP nucleotides are incorporated into the RNA segments during transcription. The products of the transcription reactions are purified (including treatment with DNAse I to remove template) using a RNeasy Kit from Qiagen.

20-100 μg of aminoallyl-modified RNA segment is coupled with NHS-ester dyes using Ambion Aminoallyl Labeling Kit. Dyes used include Alexa 488, Alexa 594 and Alexa 647 (Invitrogen/Molecular Probes) as well as Cy3 (Amersham).

Each segment is made separately in 4 colors so that each position on the scaffold can be filled with a segment in any of the four colors; thus different colors can be added at different positions to create many unique color combinations.

In this particular embodiment, adjacent segments are of different colors or there may be dark segments interspersed so that each segment is detected as an individual 'spot'. Dark segments may be used as part of the nanoreporter code.

ANNEALING OF THE RNA SEGMENTS TO THE SCAFFOLD

Segments for each position are annealed in a 2:1 ratio of segment to M13 scaffold in 1X SSPE buffer at 70°C for 2 hours. An assembled nanoreporter with labeled RNA segments is depicted in Figure 3A-3B. Figure 3A depicts a nanoreporter in which only alternate "spots" (1, 3, 5 and 7) are labeled, and Figure 3B depicts a nanoreporter in which every spot is labeled.

PREPARATION OF THE GHOST PROBE

One or more oligonucleotides comprising sequences (of, *e.g.*, 30-70 nucleotides) complementary to different regions of the target nucleic acid(s) of interest than those to which the target-specific sequences of the reporter probe are complementary, are generated. Optionally, F-tags for F-hook attachment are ligated to the 5′ end of the ghost probe using a ligator oligonucleotide that is complementary to a short sequence on the 3′ end of the F-hook as well as a short sequence on the 5′ end of the ghost probe. The sequences that are complementary to the ligator oligonucleotide are not part of the F-hook sequence or the probe sequence, but are additional nucleotides added to those oligos in order to facilitate ligation.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1014

ATTACHMENT OF F-TAGS TO THE GHOST PROBE

An F-tag (*e.g.*, an oligonucleotide having the sequence 5′-GATGGAGAC GTCTATCATCACAGC GTCTATCATCACAGC-biotin -3′, which includes 2 copies of the complement of the F-hook 5′-GCTGTGATGATAGAC -3′, followed by 9 bases of ligator sequence and is complementary to the F-hook) is attached to the 3′ end of the ghost probe to allow for (1) purification of the ghost-probe-target-reporter hybridization complex; and (2) attachment of the hybridization complex on the slide via the biotin moiety. The sequence of the ligator for attaching F-tag to the 3′ end of the ghost probe can be 5′-GTCTCCATCTTCCGACAG-3′.

Materials:

100uM F-biotin tag
100uM F ghost probe ligator
Fermentas 10X T4 DNA Ligase Buffer
1uM ghost probes
Fermentas T4 DNA Ligase

Method:

1. Pre-anneal the hook and ligator:

5uM F-biotin tag /ligator mix
5 µl [100uM] F-biotin tag
5 µl [100uM] F-ghost probe ligator
10 µl 10X T4 DNA Ligase Buffer
80 µl DEPC H_2O

Anneal on the MJ Thermocycler (95°C, 3min; 72°C, 30sec, -1°C/cycle x 68 cycles; hold at 4°C).

2. Set up the following ghost probe ligation:

300 nM anti-F2-biotin -GP
6.0 µl [1 uM] Ghost Probe
4.8 µl [5uM] anti-F2-biotin tag /ligator mix

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1015

1.52 µl 10X T4 DNA Ligase Buffer

3.68 µl DEPC H_2O

4.0 µl T4 DNA Ligase

Ligate on the MJ Thermocycler (37°C, 18hr; 65°C, 15 minutes; hold at 4°C)

3. QC the ligation on a 15% Novex TBE-Urea gel:

Prepare the following loading solutions:

Ligation	Neg Control-Ghost Probe
3.33 µl [300 nM] ligation	1 µl [1uM] ghost probe
1.67 µl DEPC H_2O	0.33 µl 10X T4 DNA Ligase Buffer
5 µl 2X Loading Buffer	3.67 µl DEPC H_2O
	5 µl 2X Loading Buffer

Neg Control- F-biotin tag /ligator mix

2 µl [0.5uM] F-biotin tag /ligator mix

0.33 µl 10X T4 DNA Ligase Buffer

2.67 µl DEPC H_2O

5 µl 2X Loading Buffer

50 bp Oligo Ladder

4 µl Ladder

6 µl 2X loading buffer

Run of a 15% Novex TBE-Urea gel at 180V for 50 minutes.

Stain with SYBR Gold for 30 minutes.

ALTERNATIVE EMBODIMENTS

Rather than covalently coupling biotin to the single-stranded F-tag, the biotinylation of the ghost probe can also be accomplished by annealing a biotinylated oligonucleotide (DNA or RNA) with a sequence complementary to the common portion of the ghost probe. Such a sequence could be the F sequence itself, or another sequence which is added to the ghost probe in addition to the F sequence. If such an additional sequence is

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1016

added, it could be from 10-100 bases long, from 1-10 copies, with the preferred configuration being a single copy from 50-100 bases long.

BIOTINYLATION OF TARGET MRNA

There are a number of commercially available kits available for the direct labeling of an mRNA sample including *Label* IT® $\mu Array_{TM}^{Biotin}$ Mirus #MIR 8010) and Biotin-Chem-Link (Roche (1 812 149). Following manufacturer's procedures biotin labeled mRNA is added to the hybridization reaction as described in Section 3d (below) with the following modifications: Since most protocols suggest the use of poly A+ mRNA, the amount of RNA used could be reduced below the 100 ng total RNA in a typical hybridization to 10 ng and possibly 1 ng. No ghost probe should be added to this reaction. F bead post-hybridization purification is no longer required. G-bead post-hybridization purification should be used to remove unhybridized biotinylated mRNA that might compete for binding to the slide. Depending on the amount of RNA used, this may or may not be required. Alternatively, total RNA could be biotinylated without the need for purification of the poly A+ fraction. In this case, the original amount of total RNA should be used (100 ng). The use of total RNA might require modifications of the manufacturer's protocol to increase labeling efficiency.

An alternative approach would be to enzymatically generate biotinylated 1st strand cDNA or biotinylated amplified RNA (aRNA) using commercially available kits and use these in place of total or mRNA. This approach would require a redesign of the reporter probes to be in the sense orientation. Both ghost probe and F-bead post-hybridization reactions would be omitted while G-bead purification would remain for removal of non-hybridized RNA.

HYBRIDIZATION OF DUAL NANOREPORTER TO TARGET

Many hybridization conditions are sufficient for achieving gene expression data. To shorten hybridization times while maintaining reasonable hybridization efficiency, several parameters can be altered: i) increasing ghost probe and reporter concentrations, ii) fragmenting of total RNA to average size range of 200-500 bp while lowering the pH of

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1017

hybridization to 6.5, iii) using more total RNA in same hybridization volume, iv) lowering hybridization volume to approximately 10 µl. Blocking reagents such as Denhardt's and ssDNA can be removed without deleterious effects on hybridization efficiency or cross hybridization to mRNAs from different species.

The following protocol has been performed successfully with multiplexing from 1 to >500 nanoreporters with ghost probes (an example demonstrating a nanoreporter assay utilizing 25 nanoreporters is described in Example 6 above, and another example demonstrating a nanoreporter assay utilizing 509 nanoreporters is described in Example 9 below). The final concentration of all nanoreporters varies depending on 1) the concentration of each reporter and 2) the number of genes being multiplexed.

Typical total nanoreporter concentrations range from 40pM (1 gene @ 40pM) to 20nM (500 genes @40pM). Ghost probe concentrations also vary from 200pM (1 gene @ 200pM) to 100nM (500 genes @ 200pM). The example that follows describes a single multiplexed hybridization containing approximately 500 endogenous genes with positive and negative controls. Add, 11.1 µl of 2.7X hybridization mix [13.5X SSPE pH 7.5 (USB #75890), 0.27 µg/µl sheared salmon sperm DNA (Sigma #D-7656), 0.27% Tween 20 (Sigma #P-1379), and 13.5X Denhardt's reagent (Sigma D-2532)], 5 µl of gene Nanoreporter mix (0.24nM each or 123nM total, includes 509 endogenous genes and 8 hybridization controls), 4.6 µl 513 gene ghost probe mix (1.3nM each or 667 nM total, includes 509 endogenous genes and 8 hybridization controls), 1 µl of purification control reporter mix (0.5pM), 1 µl of total cellular RNA (100ng/ul), 1 µl of 30X spike target mix (1.5nM-3fM) and 6.3 µl of DEPC treated water (Ambion #9922) to a 0.2ml thin wall tube (final volume 30 µl).

Final concentration of hybridization reagents should be 5X SSPE, 0.1% Tween 20, 100ng/µl sheared salmon sperm DNA, 5X Denhardt's reagent, 40 pM each Nanoreporter (~20nM total), 200pM each ghost probe (~100nM total) and 33ng/µl of total cellular RNA. Control spike targets typically vary in range from 50 fM down to 0.1 fM in a single reaction. All reagents are most preferably free of all nuclease activity. For optimal results, all reagents should be free of nuclease activity.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1018

Mix reagents well and incubate in temperature block with heated lid for 20 hours. After hybridization purify the nanoreporters with affinity reagents for both the ghost probe and the reporter probe.

ALTERNATIVE EMBODIMENT: HYBRIDIZATION PROTOCOL FOR WITHOUT SSDNA AND DENHARDT'S REAGENT

This protocol has been performed successfully with multiplexing from 1-500 nanoreporters and ghost probes. Removal of ssDNA and Denhardt's reagent from hybridizations performed with human reagents (Nanoreporters and ghost probes) had no effect on cross hybridization with mouse total RNA when compared to a hybridization containing ssDNA and Denhardt's. In addition, removal of ssDNA and Denhardt's does not result in an increased background signal (based on negative hybridization controls). Finally, there is no significant loss (or gain) of signal for endogenous genes hybridized in the presence or absence of ssDNA and Denhardt's (509 genes, R^2 value=0.998).

ALTERNATIVE EMBODIMENT: HYBRIDIZATION CONDITIONS FOR FRAGMENTED CELLULAR MRNA

Fragmentation of cellular RNA has been achieved by both thermal and cation catalyzed protocols. These protocols were designed to obtain fragment lengths between 100 and 700 bp (on average). Thermal fragmentation: Dilute total RNA sample to 200 ng/μl in RNAse free water. Heat sample to 95°C in temperature block with heated lid.

Stop fragmentation by placing sample on ice. Use immediately or store at -80°C until use. Fragmentation via cation catalyzed reaction modified from manufacturer's protocol (Ambion). Bring volume of RNA sample up to 9 μl with RNAse free water. Final concentration of total RNA should be between 0.2 and 2 μg/ml. Add 1 μ1 of 10X fragmentation buffer (Ambion 10X fragmentation buffer). Incubate at 70°C for 5 minutes in temperature block. Longer times will result in smaller fragment size on average. Stop reaction by addition of 1 μl 200 mM EDTA. Use immediately or store at -80°C until use.

Fragmented RNA samples are hybridized as described herein except for the following modifications: i) pH of SSPE is reduced to 6.5 and ii) the time of reaction is

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1019

reduced to 6 hours (for hybridization reactions in which reporter probe and ghost probe concentrations are 200 pM).

PURIFICATION OF NANOREPORTER-TARGET COMPLEXES

Post-hybridization purification is preferred when the total reporter probe concentration is above 1 nM. Purification significantly decreases non-specific binding and increases specific binding efficiency to the slide at higher reporter and ghost probe concentrations. In the example provided above, a single F-bead purification is described (purifies hybridized complexes from the ghost-probe end). As described in Example 9 below, optimal results at high ghost probe concentrations (>5nM total) are obtained via a subsequent G-bead purification which purifies the hybridization complexes from the 5′ end of the reporter effectively removing excess non-hybridized ghost probes. The preferred order of purification is F-bead, then G-bead but the order can be reversed and the protocols optimized accordingly. The exact sequences used in these affinity purifications can likely be changed and optimized in alternative embodiments of the technology. These affinity purification steps and reagents are currently nucleic acid based but could theoretically be any sort of binding pairs that exhibit specific binding to one another and can be released by chemical treatment or alteration of binding conditions such that the interaction is disrupted and released. For example, an antibody/antigen pair, a protein/metal interaction, or ligand/receptor interaction, *etc.*

One example of purification is provided below.

After hybridization is complete, the salt of a hybridization sample (30 µl, starting at 5x SSPE = 825mM Na^+) is adjusted to a final concentration of approximately 1x SSPE. The diluted sample is added to 30 µl F-hook MyOne Dynabeads (F-MODB) and bound for 15 minutes at room temperature while rotating. The beads are sequestered with a magnet and the supernatant removed. The beads are washed twice with 150 µl 0.1x SSPE + 0.1% Tween at room temperature for 15 minutes with rotation and discarded. The purified reporters are eluted in 30 µl 0.1x SSPE at 45°C for 15 minutes with rotation. At this point the hybridized reporters are purified from the contaminating un-hybridized reporters. The elution still contains contaminating un-hybridized ghost probes which will compete with the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1020

reporters for biotin-binding sites on the streptavidin coated slide. The 30 µl is added to 130 µl of 1x SSPE + 0.1% Tween to increase salt concentration. The sample (150 µl) is then loaded onto 30 µl of G-MODB and bound for 15 min at room temperature. The supernatant is discarded and the beads washed with 150 µl 0.1x SSPE + 0.1% Tween at room temperature for 15 minutes with rotation. The wash is discarded and the fully purified reporters eluted with 25 µl 0.1x SSPE at 45°C for 15 minutes with rotation. At this point only targets molecules that are hybridized to both a ghost probe (containing the anti-F sequence) and a reporter (containing the anti-G sequence) will remain in solution.

IMMOBILIZATION AND STRETCHING AND IMAGING OF NANOREPORTER-TARGET COMPLEXES

Attachment to the slide and immobilization of the stretched complex may be achieved via a biotin-streptavidin interaction. In alternative embodiments, immobilization and stretching are achieved with other interaction pairs provided one of the two could be immobilized on the slide and the other attached to either the ghost probe or the reporter. Stretching does not have to be achieved via electrophoresis but can be done mechanically. The addition of bis-tris propane to the sample before binding is not required. The technology is not limited to the use of particular label monomers exemplified herein as long as the different label monomers can be separated by image processing.

One example of an immobilization and stretching protocol is provided below.

After purification, the hybridization products are loaded directly into an open well of a microfluidic device. The liquid is pulled into a microfluidic channel by capillary action where the hybridized molecules bind to the streptavidin-coated slide through the biotinylated ghost probe. The microfluidic device then intermittently tilts along the axis perpendicular to the length of the channels in alternating directions in order to force the reaction mixture to repeatedly pass through the channel and increase the binding efficiency.

After binding the hybridization reaction, the channel is washed with 1x TAE for 5 minutes by tilting the device at an angle. Fresh TAE is then added to each well to a level sufficient to contact platinum electrodes which are inserted in the wells (30 microliters in our current geometry). An electrical potential of 200V is then applied between the two

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1021

wells connected by the microfluidic channel, stretching the reporters. After one minute of pre-electrophoresis to remove any remaining contaminating un-bound reporter molecules in the channel, a solution of 0.5 μM G-hooks in 1x TAE is added to the cathodic well (60 microliters of this solution). The electrical potential draws the G-hooks through the channel toward the anodic well. As they pass through the channel, the hooks hybridize with the free G-tag sequences on the free-end of the reporters which are bound to the surface and stretched. The streptavidin on the surface then binds the biotin on the G-hook and immobilizes the free end. When the potential is removed, the reporters remain stretched for imaging.

EXAMPLE 9: HYBRIDIZATION OF 509 CELLULAR GENES TO 100NG TOTAL RNA FROM A549 CELLS USING NANOSTRING REPORTER SYSTEM.

HYBRIDIZATION REACTION

Detection of 509 endogenous cellular genes was carried out in single multiplexed hybridization reaction. Eight non-human control sequences were spiked into each reaction that corresponded to approximately 0.1, 0.5, 1, 5, 10, 50, and 100 copies per cell as well as two reporters with no target (negative controls). There were also 4 reporters added that served as positive (3) and negative (1) controls for the post-hybridization purification process. A set of negative control hybridization was also performed containing the entire Nanostring reporter library but lacking cellular RNA.

Each sample was hybridized in triplicate. Final concentrations of the hybridization reagents were as follows: 20.8 nM total Nanoreporters (521 individual Nanoreporters at 40pM each), 103 nM total ghost probe (517 individual ghost probes @ 200pM each), 5X SSPE (pH 7.5), 5X Denhardt's reagent, 100ng/ul sheared salmon sperm DNA, 0.1% Tween 20, 50fM S11 spike target DNA, 10fM S10 spike target DNA, 5fM S9 spike target DNA, 1fM S8 spike target DNA, 0.5 fM S7 spike target DNA, 0.1fM S6 spike target DNA. S3 and S4 were added as negative controls. RNA was obtained from A549 lung epithelial cells under two different conditions. The final concentration of total RNA per hybridization was 33ng/ul. No total RNA was added to the negative control hybridizations.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1022

The final volume of the reaction was 30u1. Reagents were mixed and incubated at 65°C in thermocycler block with heated lid for 20 hours.

Master mix	(1 Reaction)	(9.3 reactions)
2.7 X hybridization mix*	11.1 µl	103.2 µl
513 endogenous gene reporters (0.24nM each)	5 µl	46.5 µl
5 513 endogenous gene ghost probes (1.3nM each)	4.6 µl	42.9 µl
Purification Control reporters (0.6nM each)	1 µl	9.3 µl
30X control target mix	1 µl	9.3 µl
Total	22.7µl	211.2 µl

*Hybridization mix (13.5X SSPE, 13.5X Denhardt's reagent, 270 ng salmon sperm DNA, 0.27% Tween 20)

Reactions	1	2	3	4	5	6	7	8	9
Master mix	22.7	22.7	22.7	22.7	22.7	22.7	22.7	22.7	22.7
48.5ng/µl RNA #1	2.1	2.1	2.1	0	0	0	0	0	0
48.4ng/µl RNA #2	0	0	0	2.1	2.1	2.1	0	0	0
H_2O	5.2	5.2	5.2	5.2	5.2	5.2	7.3	7.3	7.3
Total Rxn volume	30 µl	30 µl	30 µl	30 µl	30 µl	30 µl	30 µl	30 µl	30 µl

Incubate reactions in thermocycler with heated lid overnight (20 hours).

POST-HYBRIDIZATION PURIFICATION

Hybridization reactions were purified to remove unhybridized reporters using an oligonucleotide complimentary to ghost probe attached to magnetic beads (F-bead). Hybridization reactions were diluted 5 fold in 0.1% Tween-20/TE to bring the final salt concentration to 1X SSPE. The diluted hybridization solution was then added to 100ul of F-beads (in 0.1% Tween-20) and allowed to bind to the beads at room temperature for 30min with continuous rotation. The beads were then washed three times in 150ul of 0.1X SSPE/0.1% Tween-20 and eluted in 100ul of 0.1X SSPE/0.1% Tween-20 for 15min at 45°C.

After F-bead elution, samples were purified from the opposite end of the hybridized complex using G-beads. Elutions were brought to a final concentration of 1X

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1023

SSPE by the addition of 50ul of 3X SSPE/0.1% Tween-20 and bound to 30ul of G-beads (in 0.1% Tween-20) for 15min at room temperature with rotation. Beads were then washed as above and eluted in 30u1 of 0.1X SSPE/Tween-20 and prepared for binding as described below.

BINDING, STRETCHING, AND IMMOBILIZATION

The samples were prepared for binding by addition of 1 ul of 1/5000 dilution of 0.1 uM Tetraspec™ fluorescent microspheres (product # T7279, Molecular Probes). Samples were loaded into a Nanostring fluidic device and attached to Accelr8 Optichem® slide coated with streptavidin (product #TB0200) by tilting the device 45deg for 15min and repeated a total of 4 times. After loading, slide surface was washed once with 90ul of 1X TAE. After wash buffer is removed the sample is prepared for electrostretching by addition of 40u1 of TAE to each well. Attached complexes were stretched by applying 200V across the fluidic channel. After 1 minute the samples were immobilized in the stretched position by adding 60u1 of 500nM of G-hook oligo solution to the well containing the negatively charged electrode while continuing to apply voltage for 5 minutes. After immobilization the TAE solution is removed and replaced with anti-photobleaching reagent for imaging.

IMAGING

Slides were imaged on Nikon Eclipse TE2000E equipped with a metal halide light source (X-cite 120, Exfo Corporation) and a 60X oil immersion lens (1.4 NA Plan Apo VC, Nikon). For each field of view, 4 images at different excitation wavelengths (480, 545, 580 and 622) were acquired with an Orca Ag CCD camera (Hamamatsu) under control of either Metamorph (Universal Imaging Corporation) or custom software. Images were processed with custom image processing software.

DATA ANALYSIS

Raw data was extracted from processed images using custom software. Data was normalized to the average counts for control spikes in each sample. To determine if a gene was "detected" by the system, the counts obtained for each gene from hybridizations

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1024

containing RNA were compared to average counts of the two negative controls using a Student's t-test. The number of genes detected was 441 (87%) and 445 (88%) in sample #1 and #2, respectively.

A scatter plot (Figure shows normalized and average $\log_2$ signal values from each positive sample (n=3) for all 509 genes. The genes that were significantly different in the two samples were identified by a T-test of signal values in sample #2 against sample #1. In the graph below, the solid lines indicate the 2-fold upregulated threshold (black line) and 2-fold downregulated threshold (gray line) relative to sample #1. Genes with significant fold changes (p-value < 0.05) are shown in solid black diamonds. Genes whose fold change p-values were above this threshold are shown in open black squares.

EXAMPLE 10: DETECTION OF SMALL SPOTS

As mentioned above, the label attachment regions of a nanoreporter scaffold region have a length anywhere from 10 nm to 10,000 nm, but preferably corresponds closely to the smallest spot that can be detected with standard optics, which is about 300 nm. Spots of different color (spectrally distinguishable) are spatially resolvable at closer spacing than spots of the same color. It is possible to fit one, two, three or four spots of different colors between two spots of the same color, and yet spectrally and spatially resolve all the spots. It is also possible to significantly reduce the distance between two spots of the same color.

The limits of spatial resolution, *i.e.*, differentiating closely spaced spots of the same color, are often thought of as hard limits, *i.e.*, the Rayleigh Criteria (Inoue, S., Spring, Video Microscopy (Plenum Press, 1997), p 30). There are many techniques to drive beyond these limits that involve different imaging and/or image processing techniques. On the imaging side, structured illumination is one method to resolve spots of the same color that are spaced closer together. 50 nm has been demonstrated but, in theory, resolution with structured illumination is unlimited (Gustafsson, 2005, Proc. Nat'l. Acad. Sci. U.S.A. 102:13081-13086). On the image processing side, mixture modeling is an effective technique to push beyond commonly accepted limits (Thomann *et al.*, 2002, J. Microsc. 211:230-248). The combination of these techniques allows for drastically smaller nanoreporters with smaller spots, corresponding to label attachment regions of less than 50 nm.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1025

These smaller spot spacings could allow for drastically shorter and more stable reporters, a larger number of codes, as well as a higher degree of multiplexing before the entanglement threshold is passed (for an explanation of entanglement thresholds, see Example 9 (described in Section 14) above.

The tradeoff of making the spots much smaller and the reporters much shorter would be decreased signal and slower scan times. However, other technical advances, such as brighter light sources, and more efficient CCDs may offset the increased scan times making these approaches reasonable.

EXAMPLE 11: COMPARISON OF NCOUNTER GENE EXPRESSION SYSTEM WITH MICROARRAYS AND TAQMAN® PCR

In one embodiment, the present invention provides a novel technology to capture and count specific nucleic acid molecules in a complex mixture. This system can be used to detect any type of nucleic acid in solution and, with appropriate recognition probes, can be modified to detect other biological molecules as well. In this Example, we focused on mRNA expression profiling. In brief, a multiplexed probe library was made with two sequence-specific probes for each gene of interest. The first probe, which we refer to as a capture probe (Figure 22a), contained a 35 to 50 base sequence complementary to a particular target mRNA plus a short common sequence coupled to an affinity tag such as biotin. The second probe, which we refer to as the reporter probe, contained a second 35 to 50 base sequence complementary to the target mRNA that was coupled to a color-coded tag that provides the detection signal. The tag consisted of a single-stranded DNA molecule, which we refer to as the backbone, annealed to a series of complementary *in vitro* transcribed RNA segments each labeled with a specific fluorophore (Figure 22a). The linear order of these differently-colored RNA segments created a unique code for each gene of interest.

To detect transcripts, unique pairs of capture and reporter probes were constructed for each gene of interest. All probes were mixed together with total RNA in a single hybridization reaction that proceeds in solution. Hybridization results in the formation of tripartite structures, each comprised of a target mRNA bound to its specific reporter and capture probes (Figure 22a). Unhybridized reporter and capture probes were removed via

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1026

affinity-purification, and the remaining complexes were washed across a surface that was coated with the appropriate capture reagent (e.g. streptavidin). After capture on the surface, an electric field was applied to the solution which extended and oriented each complex in the same direction. The complexes were then immobilized in an elongated state (Figure 22b), and imaged (Figure 22c). Each target molecule of interest was identified by the color code generated by the ordered fluorescent segments present on the reporter probe. The level of expression was measured by counting the number of codes for each mRNA.

In this work, we demonstrated the linearity, reproducibility, and sensitivity of the nCounter system of the present invention and demonstrated that fold-change measurements of significantly regulated genes correlated well with microarrays, and even better with real-time PCR. In addition, we showed that the nCounter system can detect low abundance mRNAs that are declared "Absent" by DNA microarrays. The validity of this detection was confirmed for a subset of genes using real-time PCR. These results demonstrate the advantages of the methods and systems of the present invention and demonstrate that they can fill an immediate niche in the expression analysis of hundreds of genes across many samples. Applications include translational medical studies, research involving gene regulatory systems, diagnostic fingerprinting, and validation of high-throughput gene expression experiments.

NCOUNTER GENE EXPRESSION SYSTEM OVERVIEW

The basis of the nCounter system is the unique code assigned to each gene to be assayed. As outlined below under Methods, we used 7 positions (visualized as "spots") and 4 colors. The 4 colors were chosen to minimize spectral overlap during imaging. The number of positions was based on a combination of factors that include the length of the DNA backbone, the minimum spot size that can be resolved under current imaging conditions, flexibility in code selection for modestly-sized gene sets (i.e. <1000 genes) and the number of potential codes for future versions of the system (4^7=16,384 if all possible combinations of codes are used). The total number of codes required for the experiments described below was 524 (15 controls and 509 genes) or roughly 3% of the available codes in a seven-spot system.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1027

Specific reporter and capture probes were synthesized in 96-well plates using a semi-automated process (see Methods). Briefly, gene-specific probes were ligated to reporter backbones, and each ligated backbone was annealed to a unique pool of seven dye-coupled RNA segments corresponding to a single code. The reporter probes were then pooled and purified using a common sequence at the end of each backbone (the 5'-repeat sequence, see Fig. 22a) to remove excess probe oligonucleotides and dye-coupled RNA segments. Capture probes were made by ligating a second sequence-specific oligonucleotide for each gene to a universal sequence containing biotin (see Fig 22a). After ligation, the capture probes were also pooled and affinity-purified using the universal sequence to remove the excess unligated gene-specific oligonucleotides. Reporter and capture probes were combined into a single "library" and used as a single reagent in subsequent hybridizations.

The expression levels of all selected mRNAs was measured in a single multiplexed hybridization reaction. The sample was combined with the probe library, and hybridization occured in solution. After hybridization, the tripartite hybridized complexes (Fig. 22a) were purified in a two-step procedure using magnetic beads linked to oligonucleotides complementary to universal sequences present on the capture and reporter probes (see Methods). This dual purification process allowed the hybridization reaction to be driven to completion with a large excess of gene-specific probes, as they were ultimately removed and thus did not interfere with binding and imaging of the sample. All post hybridization steps were handled robotically on a custom liquid-handling robot (Prep Station, NanoString Technologies). The Prep Station can process 12 samples in 2.5 hours for a total of 48 assays per instrument in 10 hours.

Purified reactions were deposited by the Prep Station into individual flow cells of a sample cartridge, bound to a streptavidin-coated surface via the capture probe, electrophoresed to elongate the reporter probes, and immobilized (see Fig 22). After processing, the sample cartridge was transferred to a fully automated imaging and data collection device (Digital Analyzer, NanoString Technlogies). The expression level of a gene was measured by imaging each sample in 4 colors and counting the number of times the code for that gene is detected. For each sample, over 600 fields-of-view (FOV) were imaged (1376 X 1024 pixels) representing approximately $10mm^2$ of the binding surface. Typical imaging

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1028

density was 100-200 counted reporters per field of view depending on the degree of multiplexing, the amount of RNA, and overall gene expression levels. However the system is capable of operating at densities 5-10 fold higher. The Digital Analyzer can accommodate up to 6 cartridges at once and current scan times for 600 FOV were 4 hours per sample cartridge. Unattended, it can process 72 samples in 24-hours per instrument.

Image processing and code counting was performed (see Methods). To minimize false positives, a reporter must meet stringent criteria concerning the number, size, brightness and spacing of the spots to ensure that the code is interpreted correctly. Reporters that did not meet all of these criteria were discarded. Using these criteria, approximately 20% of the detected molecules were counted. No parity schemes or error correction were employed in the current system. Data was output in simple spreadsheet format listing the number of counts per gene per sample.

EXPERIMENTAL DESIGN

To demonstrate the utility of the NanoString nCounter system, we performed a series of experiments in which the expression levels of 509 genes were assayed with NanoString's nCounter system. 347 of these genes were selected from previous microarray studies of poliovirus (PV)-infected A549 cells and the remaining 162 genes were a selection of previously-designed probes added to bring the multiplex total to over 500. Additional experiments with other probe libraries were performed with commercially-available RNAs and total RNA isolated from developing sea urchin embryos. We compared the nCounter results to those obtained with the Affymetrix GeneChip sytem and with real-time PCR measuring the same total RNA samples.

Table 5 summarizes the results obtained using a set of 14 genes tested on all three platforms. They are listed by RefSeq Accession numbers, Probeset ID, and TaqMan ID. Signal levels for both samples in all three platforms are shown with standard deviations in parentheses. Values shown correspond to normalized counts for the nCounter system, RMA normalized intensity for Affymetrix's GeneChip, and cycle threshold (Ct) for ABI TaqMan assay. Detected (D) and Undetected (U) calls are based on platform-specific criteria. For the Affymetrix platform, a gene was only considered undetected if all 3 replicates for each

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1029

sample were called "Absent" by the MAS 5 algorithm. All genes were detected by the TaqMan assay based on a cutoff of less than 35 cycles. Fold-change comparisons are shown in Figure 26b.

Table 5. Comparison of signal levels and detected/undetected calls for 14 genes on the nCounter, GeneChip and TaqMan platforms.

Accession#	Gene Name	Affymetrix Probeset ID	TaqMan ID	NanoString			Affymetrix			TaqMan		
				Mock	PV	Mock/PV	Mock	PV	Mock/PV	Mock	PV	Mock/PV
				signal		Detection	signal		Detection	Ct		Detection
NM_005570	*LMAN1*	203293_s_at	Hs00194366_m1	669 (63)	224 (10)	D/D	61 (11)	45 (9)	D/U	25.5 (0.17)	27.2 (0.11)	D/D
NM_020726	*NLN*	225943_at	Hs00252959_m1	428 (78)	164 (10)	D/D	545 (54)	443 (47)	D/D	25.7 (0.03)	27.4 (0.09)	D/D
NM_015884	*MBTPS2*	206473_at	Hs00210639_m1	347 (37)	111 (9)	D/D	48 (3)	48 (8)	U/U	26.7 (0.04)	28.5 (0.09)	D/D
NM_002895	*RBL1*	1555004_a_at	Hs00161234_m1	270 (39)	108 (7)	D/D	62 (8)	51 (3)	D/D	27.4 (0.05)	28.8 (0.10)	D/D
NM_006219	*PIK3CB*	217620_s_at	Hs00178872_m1	204 (31)	73 (9)	D/D	23 (0)	24 (3)	U/U	28.0 (0.07)	29.5 (0.15)	D/D
NM_016436	*PHF20*	209423_s_at	Hs00363134_m1	195 (19)	70 (8)	D/D	47 (4)	54 (9)	U/U	27.9 (0.02)	28.8 (0.05)	D/D
NM_014484	*MOCS3*	206141_at	Hs00819330_s1	183 (6)	83 (9)	D/D	42 (2)	40 (2)	D/U	28.5 (0.15)	29.1 (0.30)	D/D
NM_025209	*EPC1*	223875_s_at	Hs00228677_m1	111 (22)	57 (6)	D/D	30 (1)	31 (2)	U/U	27.2 (0.07)	28.9 (0.07)	D/D
NM_018094	*GSPT2*	205541_s_at	Hs00250696_s1	100 (43)	76 (3)	D/D	214 (12)	102 (18)	D/D	30.3 (0.22)	30.7 (0.04)	D/D
NM_006420	*ARFGEF2*	215931_s_at	Hs00197455_m1	77 (2)	29 (1)	D/D	42 (5)	47 (6)	U/U	26.6 (0.12)	28.3 (0.23)	D/D
NM_007211	*RASSF8*	207754_at	Hs00200537_m1	62 (13)	31 (7)	D/D	37 (2)	37 (2)	D/U	27.3 (0.05)	28.5 (0.09)	D/D
NM_020800	*IFT80*	226098_at	Hs00398803_m1	41 (6)	29 (5)	D/D	321 (15)	123 (22)	D/D	29.0 (0.06)	29.8 (0.41)	D/D
NM_015139	*SLC35D1*	209713_s_at	Hs00209446_m1	38 (1)	20 (3)	D/D	42 (7)	43 (0)	U/U	27.8 (0.03)	29.1 (0.16)	D/D
NM_153034	*ZNF488*	229901_at	Hs00399237_m1	31 (8)	13 (4)	D/U	114 (7)	92 (15)	D/D	29.1 (0.09)	30.0 (0.80)	D/D

METHODS

Cell culture; infection; and RNA isolation

A549 cells, a human lung epithelial cell line, were purchased from ATCC. Poliovirus (PV) stocks were the kind gift of Kurt Gustin's laboratory (University of Idaho). Sub-confluent A549 cells were either mock-infected or infected with PV at a multiplicity of infection of 50. Virus was adsorbed for 30 minutes at 32°C in PBS supplemented with 10mM $MgCl_2$ and 10mM $CaCl_2$. Following adsorption, residual virus was removed and DMEM with 10% FBS, 2mM L-Glutamine and Penicillin-Streptomycin was added. After 5 hours of

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1030

infection, the total RNA was extracted using Qiagen RNeasy mini-spin columns according to the manufacturer's protocols. Two independent mock- and PV-infections were performed. Following RNA isolation, the RNA from the replicates was pooled to create one sample of RNA from PV-infected cells and another from mock-infected cells. Aliquots of these two RNAs were used in all subsequent microarray, real-time PCR and nCounter analyses.

Control Target Preparation

Targets for spike-in controls consisted of 100-base HPLC purified oligonucleotides that were complementary to the spike-in reporter and capture probes. These and all other oligonucleotides were purchased from Integrated DNA Technologies. They were generated to specific 100-base regions of the following non-human sequences and arbitrarily named A-H [spikes A, E and F, (accession number AY058658.1); spikes B-D, (accession number AY058560.1), and spikes G and H, accession number DQ412624)].

Generation of fluorescent RNA segments

To prepare the RNA segments for reporter probe synthesis, PCR fragments for each segment were generated using primers specific to M13 and containing either T7, T3 or SP6 RNA polymerase promoters. RNA transcripts were *in vitro* transcribed from these templates using the Megascript kit (Ambion) in the presence of 50% amino-allyl UTP (Sigma). Each of the seven resulting amino-allyl labeled RNA transcripts was coupled to one of 4 NHS-ester fluorophores [Alexa 488, Alexa 594, Alexa 647 (Invitrogen) or Cy3 (GE Healthcare)].

NanoString reporter preparation

NanoString reporters consisted of linearized single-stranded M13 DNA, referred to as backbone, annealed to fluorescently-labeled, *in vitro* transcribed RNA segments. Using standard molecular biology protocols, circular single-stranded M13 (United States Biological) was linearized, and an oligonucleotide containing four 15-base repeats, referred to as the 5'-repeat, was ligated on to the 5' end of the backbone. Using a Hamilton STAR liquid-handling robot, a master mix containing a universal oligonucleotide that served as a ligation "bridge" plus ligase buffer was added to individual wells of 96-well plates containing normalized (10μM) gene-specific oligonucleotide probes (35-50 bases). After a short incubation at 37°C to anneal the probe oligonucleotide to the complementary portion of

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1031

the bridge oligonucleotide, ligation was initialized by addition of another master mix containing the equivalent of 1.2 pmoles of M13 backbone per well, additional ligation buffer, and T4 ligase. Plates were incubated at 37°C in a 96-well thermocycler for 2h. The efficiency of the ligation reactions was assessed by cutting the backbone approximately 600 bases away from the ligation site using short oligonucleotides to generate double-stranded restriction sites, and analyzing the size of the resulting fragments by PAGE. Ligation reactions were desalted via centrifugation through G-50 Sephadex columns in a 96-well format.

Each gene-specific backbone was assigned a unique code consisting of an ordered series of differently-colored RNA segments annealed to the backbone. Sets of seven approximately 900-base fluorescently-labeled RNA transcripts complementary to distinct sequences on the backbone were created in 96-well plates using a Hamilton STAR robot. Each well received a unique combination of RNA segments that, when annealed to the M13 backbone and visualized in linear sequence, resulted in a unique code. Plates containing RNA segment pools were mixed with probe-ligated M13 backbones in a 2:1 molar ratio. Annealing of segments to the backbone was performed in individual wells of a 96-well PCR plate. At the same time, one unlabeled RNA segment was also annealed to each reporter to cover the remaining single-stranded region of the backbone, leaving only the probe at one end and the 5'-repeat at the other as single-stranded DNA. The rest of the reporter is a double-stranded DNA/RNA hybrid. To remove excess RNA transcripts and unligated probes, the reporters were then pooled and affinity-purified over magnetic beads (Dynal, Invitrogen) coupled to oligonucleotides complementary to the 5'-repeat sequence on the 5' end of each backbone. The final reporter molecules had seven labeled regions in a linear sequence each of which resulted in a ~300nm spot when imaged by an epi-fluorescent microscope under the conditions described below.

NanoString capture probe preparation

The capture probe consisted of a 35- to 50-base gene-specific sequence attached to a capture-oligonucleotide comprised of two 15-base repeats, referred to as 3'-repeats, linked to a biotin molecule. In a process similar to reporter probe synthesis, normalized gene-specific oliogonucleotides were annealed to a short universal "bridge" oligonucleotide in ligation buffer. A master mix containing the 3'-repeat oligonucleotide,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1032

additional ligation buffer, and T4 ligase was added. The 3'-repeat oligonucleotide was present in 4-fold excess. Ligation reactions were performed in 96-well plates in a thermocycler for 2h at 37°C. The efficiency of each ligation was assessed by PAGE. After ligation there are 3 potential species of molecules in the reaction: the 3'-repeat ligated to the gene-specific probe (the "capture probe" in Fig. 22), the excess unligated 3'-repeat, and any residual unligated probe oligonucleotide if the reaction did not go to completion. Excess free probe is the only species that negatively affects the hybridization results as it competes for target with the fully-ligated capture probe. Therefore, after ligation the capture probes were pooled and purified over magnetic beads coupled to an oligonucleotide complementary to the 3'-repeat to remove free probe oligonucleotide. A later post-hybridization purification step removed excess unligated 3'-repeat oligonucleotide (see the anti-5'-repeat post-hybridization purification, below).

Probe Design and Selection

Potential pairs of 50-base probes were chosen by first screening 100-base target regions of the mRNA to eliminate long direct and inverted repeats, high GC content, and long poly-C stretches (due to the difficulty in synthesizing poly-G sequences in probe oligonucleotides). The refined list of target regions was then screened for cross-hybridization using NCBI BLAST [13] (version 2.2.14) to align them against the Human RefSeq mRNA database[1] (Hs: release 17). These 100-base target BLAST alignments were used to filter out targets that resulted in either 50-base probe having greater than 85% identity or stretches greater than 15 contiguous bases complementary to any non-target mRNA. The cross-hybridization cutoffs were chosen based on prior 50-base hybridization and probe design studies. [14, 15] Probes were then screened for inter- and intra-reporter and capture probe interactions and selected for probe pairs with calculated melting temperatures (T_m) between 78-83°C, with an ideal target of 80.5°C. In the last stage of selection, probes that met all requirements but had a calculated T_m greater than 83°C were dynamically trimmed until the T_m was calculated to be less than or equal to 83°C with a minimum-length cutoff of 35 bases. Final probe-pair selection was based on a score calculated from cross-hybridization and T_m screens, with preference given to probes which did not need to be trimmed to meet T_m requirements.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1033

NanoString reporter gene libraries

The reporter library for the A549 cell study contained probes to 509 human genes. The majority of these genes (347) were selected based on previous microarray studies on PV infected A549 cells (unpublished) using the Limma package in Bioconductor [16] to identify genes with a false detection rate of less than 0.05. The remaining 162 genes were collected from a variety of other studies; they have no particular biological relevance to the PV study, but were added to evaluate the ability of the nCounter assay to multiplex more than 500 genes. The list of 509 RefSeq mRNAs was based on the current human genome organization (HUGO) gene name associated with the list of Affymetrix probe set IDs. Note that not all of the target regions for the Affymetrix probe sets overlap completely with the RefSeq mRNAs. The reporter library for the MAQC-consortium study contained probes to 35 human genes that were selected based on the RefSeq gene list published in the MAQC consortium study.[2] The probe library for the *Strongylocentrotus purpuratus* study contained probes to 55 *S. purpuratus* genes including polyubiquitin, which was used for normalization purposes, and seven probes to *Homo sapiens* genes, which were used as the negative controls. The analysis described in this paper only includes the 21 *S. purpuratus* genes for which there was comparable real-time PCR data available. All libraries described also contained 8 non-human control probe pairs (spike-ins) and multiple control reporters that did not contain gene-specific probes, but were used to assess purification and binding efficiencies.

Hybridization Reactions

Detection of cellular transcripts was carried out in multiplexed hybridization reactions. Each sample was hybridized in triplicate with final concentrations of the hybridization reagents as follows: 200pM each capture probe, 40pM each reporter probe, 5X SSPE (pH 7.5), 5X Denhardt's reagent (Sigma), 100ng/μl sheared salmon sperm DNA (Sigma), and 0.1% Tween-20. Each 30μl hybridization reaction also contained 100ng total RNA at a final concentration of 3.3ng/μl. In addition, 6 positive and 2 negative control probe-pairs to non-human sequences were added to each reaction. Final concentrations of the 100-base control targets were 50fM spike A target, 10fM spike B target, 5fM spike C target, 1fM spike D target, 0.5fM spike E target, and 0.1fM spike F target. No target was added for

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1034

spikes G and H (negative controls). Reagents were mixed and incubated at 65°C in a thermocycler block with a heated lid for 20 hours.

Post-hybridization purification

To remove unhybridized reporters, reactions were purified over magnetic beads (Invitrogen) coupled to oligonucleotides complementary to the 3'-repeat sequence contained on every capture probe. Reactions were first diluted to 1X SSPE in 0.1% Tween-20/TE and allowed to bind to beads at 22.5°C for 30 minutes with continuous rotation. The beads were washed three times in 150μl of 0.1X SSPE/0.1% Tween-20 and the hybridized complexes eluted in 100μl of 0.1X SSPE/0.1% Tween-20 for 15 minutes at 45°C. After elution, samples were purified a second time to remove excess capture probes by binding to magnetic beads coupled to oligonucleotides complementary to the 5'-repeat sequence contained on every reporter probe. The elutions from the anti-3'-repeat beads were brought to a final concentration of 1X SSPE by addition of 50μl of 3X SSPE/0.1% Tween-20 and bound for 15 minutes at 22.5°C with rotation. Beads were washed as above and eluted in 30μl of 0.1X SSPE/0.1% Tween-20 at 45°C. The doubly-purified samples were then prepared for capture as described below.

NanoString reporter capture, stretching, and imaging

One microliter of 1/5000 dilution of a 0.1% solids solution of a custom-formulation of Tetraspeck fluorescent microspheres (Invitrogen) was added to each sample. Samples were loaded into a NanoString fluidic device made by lamination of laser-machined cast acrylic with a coverslip coated with streptavidin (Optichem®, Accelr8 Technology Corporation) using a laser-cut double-sided adhesive layer (Fralock) to generate 30μm deep microfluidic channels. The samples were driven through the channel by hydrostatic pressure and bound specifically by the biotinylated 3' end of the capture probe. After capture, the surface was washed once with 90μl of 1X TAE and prepared for stretching by the addition of 40μl of TAE to each well. Reporter probes were stretched and aligned by applying 160V/cm for 1 minute along the fluidic channel. Stretched reporters were then immobilized to the surface by addition of 60μl of a 500nM solution of a biotinylated oligonucleotide complementary to the 5'-repeats present on the 5' end of all reporter probes. The current remained on for 5 minutes, throughout the immobilization process. After immobilization, the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1035

TAE solution was removed and replaced with a custom formulation of the anti-photobleaching reagent SlowFade (Invitrogen) for imaging.

Slides were imaged on a Nikon Eclipse TE2000E equipped with Perfect Focus, a 1.4 NA Plan Apo VC 60X oil-immersion lens (Nikon), an X-cite 120 metal halide light source (Exfo Corporation), an automated H117 stage (Prior Scientific), and a SmartShutter (Sutter Instrument). For each field of view, 4 images at different excitation wavelengths (480, 545, 580 and 622) were acquired with an Orca Ag CCD camera (Hamamatsu) under control of either Metamorph (Universal Imaging Corporation) or custom software.

Image processing

Image processing was performed on 4 images (one for each wavelength) on a FOV-by-FOV basis. The custom algorithm treats each FOV as a fundamental block in which the following basic steps are performed: 1) spot identification, 2) image registration, 3) spatial clustering to produce strings, and 4) string classification.

In the first step of the algorithm, spots were identified. The background intensity level of each channel was computed and used to threshold the image into signal and background, where signal regions are the result of a specific wavelength of light observed as a point spread function (PSF). The signal mask was segmented using a custom Watershed algorithm. The segmented regions were then labeled, parameterized, and filtered to remove non-PSF spots. The remaining spots were centrally archived for use in registration and reporter calling.

Image registration was performed on each FOV based on archived spots that correspond to fluorescent beads (fiducials) that were bound to the imaging surface (see NanoString reporter capture, stretching, and imaging). The archived spots were cross-referenced to identify inter-channel clusters of spots that meet fiducial requirements (interchannel intensity thresholds and ratios). Clusters that met requirements were archived as fiducials. The final list of fiducials represented the spatial transforms that occured between channels during image acquisition. Spatial offsets were as large as 5-6 pixels. The spatial transform was solved for using the observed fiducial centroids and their pre transform

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1036

(assumed) coincident centroids ($X_2 = X_1 * T$). The inverse transform was then applied to all identified spots to restore their original centroids.

After spot identification and image registration, spots were assembled into "strings" via clustering. At this point, each string was filtered to remove any spots attributed to bleed-though signal. The filtered strings were then classified as reporters or non-reporters. To be classified as a reporter the string must contain the correct number of spots, meet specific spot-to-spot spacing thresholds (1.2-2.9 pixels), and meet acceptable linearity and orientation requirements. Clusters that were classified as reporters were then counted and summed for each gene over all FOVs.

NanoString data normalization and analysis

To account for slight differences in hybridization and purification efficiency, data was normalized to the average counts for all control spikes in each sample. To determine if a gene was "detected" by the NanoString system, the triplicate measurements obtained for each experimental gene were compared to triplicate measurements for the two negative controls. For a gene to be categorized as detected, the average counts for the experimental gene had to be greater than the average counts for the 2 negative controls, and the Student's T-test P-value had to be less than 0.05. For the *S. purpuratus* study, the data were normalized to the polyubiquitin gene and detected genes were determined by a Student's T-test against the 7 human negatives.

Production of Affymetrix array data

Aliquots of the same RNA samples analyzed by the NanoString nCounter system were also analyzed by microarray. In brief, triplicate samples of 100ng of total RNA were analyzed on Human U133 Plus 2 arrays. Since 1-2μg of total RNA is typically required for the standard Affymetrix single amplification protocol, the RNA expression data was produced following the manufacturer's standard protocol using the GeneChip® Two-Cycle Target Labeling kit (Affymetrix part #900494). Hybridization, washing and staining were carried out using the manufacturer's standard protocols. Data was normalized using RMA. Affymetrix "presence/absence" calls were obtained by independently processing the data with MAS 5.0 algorithm. The array and NanoString data have been made public via the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1037

Array Express database (E-MEXP-1072).[17] For data in Figure 25, an Affymetrix probe set was declared detected if any one of the 3 replicates was called "present" or "marginal".

TaqMan real-time PCR Data

Genes which showed discordant levels of expression between the NanoString and microarray systems were selected based on the following criteria: 1) genes had to be significantly differentially expressed in one platform (greater than 2-fold, P-value <0.05) and not in the other platform (less than 1.5-fold, P-value >0.05); 2) both the Affymetrix and NanoString probe sets had to map to the same RefSeq mRNA; and 3) an inventoried ABI TaqMan® probe set had to be available. For each sample, 4μg of total RNA was reverse-transcribed using random hexamers in a final volume of 40μl. The reactions were diluted to 200μl in TE and then 5μl, equivalent to 100ng of total RNA, was used in each real-time PCR reaction. All assays were performed in triplicate. The data were normalized to Beta-glucuronidase (*GUS*).

MAQC comparisons

A library of 35 RefSeq mRNAs that were also listed in the MAQC TaqMan real-time PCR data set[2] was used to analyze differential gene expression between the two commercially-available reference RNAs, Human Reference total RNA (Stratagene) and Human Brain Reference total RNA (Ambion). As described in the original study[2], genes that were not detected in all samples for both the NanoString and TaqMan platforms were removed from further analysis. *STAT5A* was removed from the NanoString data due to a known cross-hybridization issue with *STAT5B*. Fold-change correlation of NanoString results with MAQC Taqman real-time PCR data for the remaining 27 genes was determined by plotting the $\log_2$ ratio of normalized signal values (Human Reference RNA versus Human Brain Reference RNA) and calculating the linear correlation coefficient for that plot.

SYBR®Green real-time PCR methods

S. purpuratus total RNA isolation, cDNA synthesis and real-time PCR was carried out as described.[6, 18] Twenty-one S. purpuratus genes were assayed by quantitative real-time PCR. The primers used for PCR are available at http://sugp.caltech.edu/resources/methods/q-pcr.psp. All genes were assayed in quadruplicate.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1038

NANOSTRING NCOUNTER GENE EXPRESSION SYSTEM PERFORMANCE

Hybridization reactions were performed in triplicate with total RNA samples isolated from mock- and PV-infected A549 cells. Each reaction contained 100ng of total RNA plus reporter and capture probes for 509 human mRNAs contained in the RefSeq database.[1] In addition, 6 pairs of positive and 2 pairs of negative control reporter and capture probes were included in every reaction. The spike-in controls produced a standard concentration curve for every hybridization reaction and were used to normalize the data for slight differences in hybridization, purification and capture efficiencies.

We first examined the linearity, dynamic range, and reproducibility of the six positive controls. Figure 23a shows the results of the control measurements from each hybridization reaction with RNA from mock- and PV- infected cells (n=6). The control signal values (counts) for each replicate were very reproducible between 0.5fM and 50fM as indicated by overlapping points on the log-log plot. The assay was also highly linear over 2.5 logs of concentration with linear regression correlation coefficients of counts vs. concentration at $\geq$ 0.998 (Figure 23b).

We then examined the sampling efficiency and the lower limit of detection. The sampling efficiency of the system was estimated by dividing the number of counts for a spike-in target by the theoretical number of molecules of that target in the reaction. For example, there were a total of approximately 1800 molecules of the 0.1fM spike-in target in each reaction. The average measurement for this target in the mock sample was 25 counts, resulting in a sampling efficiency of approximately 1%. The limit of detection of the assay was determined by comparing the counts for the positive control at the lowest concentration to the counts of the negative controls using a Student's T-Test (see Methods). The lowest concentration of controls detected in the context of the 500-plex hybridization reaction was between 0.1 fM and 0.5fM in a total volume of 30µl containing 100ng of total RNA. Background signal for the two negative controls averaged 14.4 +/- 6.5 and 10.2 +/-3.5 for the mock and PV-infected cells, respectively. Assuming 10pg of total RNA/cell (i.e. 10,000 cells in 100ng), the limit of detection corresponds to between 0.2 to 1 molecules of control target per cell.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1039

The reproducibility of the nCounter system in measuring the 509 mRNAs was also examined. In Figure 24a, the normalized counts for all 509 genes from two independent hybridizations of RNA from PV-infected cells (technical replicates) are shown on a log-log scale. The data demonstrate that the NanoString system is reproducible: a linear fit to the data results in a correlation coefficient of 0.9999. The average correlation coefficient of each pair-wise combination of replicate assays was 0.9995+/- 0.0004. This was slightly higher than that obtained from the same analysis of genes on the DNA microarray (average correlation coefficient = 0.9934+/-0.0059). In addition, Figure 24a shows that endogenous genes were detected with signals ranging from about 25 counts to over 50,000 counts, which suggests that the dynamic range of the system is larger than the 2.5 logs tested with the positive spike-in controls.

An important feature of any gene expression technology is determining the relative difference in gene expression between two or more samples. We measured change in expression levels for the 509 genes in the reporter library between mock- and PV-infected cells. The results are plotted in Figure 24b (n=3). Using cutoff criteria of a 2-fold change in expression with a P-value of 0.05 or below, there were 28 genes that were induced and 115 genes that were repressed by PV infection as indicated by the upper and lower lines in Figure 24b. These results demonstrate the nCounter system can be used to measure gene expression of more than 500 genes in a single assay and identify those genes that change significantly between samples.

COMPARISONS BETWEEN NANOSTRING AND MICROARRAYS

We compared the ability of the NanoString system to detect and measure the level of endogenous transcripts against microarrays; using the widely-used Affymetrix GeneChip system as a representative microarray platform. As described above, nCounter assays were performed directly on 100ng of total RNA without amplification. The same samples and amount of RNA were also analyzed with Affymetrix U133Plus2 arrays, using the two-cycle amplification/labeling protocol recommended by the manufacturer.

In order to determine how the nCounter system compares in sensitivity to microarrays, we examined the number of genes detected in each platform. Of the 509 genes

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1040

assayed, there were 60 for which there was no acceptable corresponding Affymetrix Probe ID (based on Supplementary Table 2 of Shi et al.[2]). For the remaining 449 genes, we examined how many were called detected by each platform. The NanoString system uses a Student's T-Test of the replicate values for each gene compared to 2 negative controls (n=6) to determine the presence or absence of each gene, whereas the Affymetrix MAS 5.0 algorithm is based on the relationship between the Perfect Match and Mismatch probe sets. The average percentage of detected transcripts in both samples was higher in the NanoString assay than in the DNA microarray assay (88.4% vs. 82.6%; Figure 4a and 4b, respectively), and the boundary between detected and undetected calls was more distinct. The accuracy of the NanoString detection calls for several genes was further validated in TaqMan assays (Table 5).

The correlation of fold-change measurements for genes that change significantly in both the NanoString and Affymetrix platforms was assessed. After normalization and preprocessing of data (see Methods), the mean $\log_2$ fold-change between PV-infected and mock-infected samples was calculated for both platforms. A Student's T-Test for differential expression was performed between the samples. A threshold P-value of 0.05 without multiple testing correction was used to identify significantly regulated genes. This analysis resulted in 4 classes of genes: those that are determined to be regulated by both platforms (202 genes), by NanoString only (55 genes), or by microarray only (78 genes), and those that are not found to be regulated by either platform (114 genes). A plot of $\log_2$ ratios for all 449 genes with their significance in each platform is available in Figure 27. Figure 26a (♦) shows a comparison of $\log_2$ ratios for the 202 genes that were found to be significantly regulated in both the NanoString and microarray assays. The two platforms agree well for these 202 genes; only 4 are found to be regulated in opposite directions (dark diamonds in the upper left and lower right quadrants of Figure 26a). The correlation coefficient of a linear fit to $\log_2$ ratios between the assays was 0.788. This correlation coefficient is similar to previous results comparing different array platforms, as well as comparisons with other quantitative measurement technologies such as real-time PCR[2-4] suggesting the results can be extrapolated to other microarray platforms.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1041

TAQMAN® ANALYSIS OF SELECTED GENES

As mentioned above, there were a number a genes in which the measured $\log_2$ fold-change was significant in one platform but not the other. We selected a subset of 14 of these genes for further analysis by TaqMan real-time PCR. Selection criteria are described in Methods. Twelve genes were determined to be differentially expressed by the NanoString assay and two by the microarray assay. TaqMan real-time PCR was performed using RNA from the same master stock of mock- and PV-infected samples, and $\log_2$ fold changes were calculated. Overall, the NanoString assay showed much higher concordance with the TaqMan assay than did the DNA microarray assay (Figure 26b). Nine of the 12 genes met the same fold change criteria by real-time PCR and the other 3 showed similar trends but had slightly higher p-value (*ZNF488*) or missed the 2-fold cutoff criteria (*MOCS3* and *PHF20*). In contrast, neither of the two genes determined to be regulated by the Affymetrix system alone (*GSPT2* and *IFT80*) were validated by the TaqMan assay.

Using the same set of 14 genes, we also compared the sensitivity of each platform by its ability to detect each gene in the two samples (Table 5). All 14 genes were detected in both samples by real-time PCR in less than 35 cycles. The results were similar for the NanoString system, with 13 of the 14 samples being detected in both samples and 1 gene (*ZNF488*) detected in mock-infected but not the PV-infected sample. In contrast, 6 genes were declared absent in both samples by microarrays and another 3 genes were declared absent in PV-infected cells. Hence, in these experiments the sensitivity of the NanoString system was superior to that of microarrays and similar to that of real-time PCR.

COMPARISON OF NCOUNTER SYSTEM WITH MAQC DATA SET

Recently, a series of studies performed by members of the MAQC consortium utilized commercially-available reference RNA samples to compare the performance of different microarray platforms[2, 4] as well as several quantitative gene expression technologies,[5] using TaqMan real-time PCR as the benchmark technology.

An nCounter probe library was constructed that was specific for 35 RefSeq mRNAs that overlapped with the MAQC gene set. The library was hybridized to Human Reference RNA and Human Brain Reference RNA samples used by the MAQC consortium

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1042

to determine $\log_2$ fold-change values. After eliminating genes declared absent in either sample by either the nCounter or the TaqMan data (as described in Shi et al. [2]), we compared the $\log_2$ fold-change values for the remaining 27 genes. As Figure 26c shows, there was excellent correlation between the NanoString and TaqMan platforms (R^2=0.945). A similar analysis of Affymetrix microarray data (site 1, Affymetrix Inc.[2]) from the same study revealed a significantly lower correlation of R^2=0.832 for the 18 genes that met the same criteria (Figure 26c).

COMPARISON OF NCOUNTER SYSTEM AND SYBR®GREEN REAL-TIME PCR

In order to further demonstrate the sensitivity, accuracy, and dynamic range of the nCounter system, we compared it to real-time PCR in a different biological system. Total RNA was isolated from sea urchin embryos at seven time points of development (egg-70h) and either analyzed directly with the nCounter system or converted into cDNA and analyzed by real-time PCR. The transcript levels of 21 genes were examined at each time point. For the nCounter assay, all genes were combined in one library and analyzed in a multiplexed reaction. Each hybridization was performed in triplicate on 100ng of total RNA (21 assays). For real-time PCR, each gene was assayed individually in quadruplicate for each time point from 2.8ng of starting material (588 assays). For both assays, the data was normalized to ubiquitin[6].

A remarkable correlation in the relative expression patterns was observed between nCounter and real-time PCR data across the time course for all 21 genes (Figure 28). The correlation was consistent for genes that were expressed at both low (*e.g. Snail*, *Pmar 1*) and high (*e.g. Est*, *Dri*) transcript levels per embryo as well as those whose expression levels changed over 3 logs during the timecourse (*e.g. Tgif*, *Msp130*). These results confirm that the nCoutner system is capable of producing real-time PCR quality data without enzymatic or signal amplification.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1043

DISCUSSION

As demonstrated above, the gene expression analysis system described herein (nCounter) is extremely sensitive (0.1-0.5fM detection limit), reproducible (replicates averaging R^2 of 0.999 over a 3-log dynamic range), and simple to use. We have demonstrated that the nCounter system is capable of a high degree of multiplexing, measuring over 500 genes in a single reaction starting with just 100ng of total RNA sample. The overall performance of the nCounter Gene Expression System correlated well with both microarrays (R^2=0.79 over 202 genes) and real-time PCR (R^2=0.95 in MAQC) in head-to-head comparisons with the same total RNA samples. In addition, our data indicates that the nCounter gene expression system is more sensitive than microarrays and similar in sensitivity and accuracy to real-time PCR (Table 5).

The nCounter system has distinct advantages not found in the major existing gene expression technologies. First, the sample RNA is measured directly without amplification or cloning. Thus, no gene-specific or 3' biases are introduced, and the levels of each transcript within a sample can be established by counting the number of molecules of each sequence type and calculating concentration with reference to internal standards. In contrast, in real-time PCR transcript concentration is calculated from the number of enzymatic steps required to attain a threshold level of product. Secondly, both the probe and target are in solution rather than bound to a surface. The reaction is driven to completion (data not shown), allowing for a higher level of sensitivity than microarrays across many target genes with lower amounts of starting material. Thirdly, NanoString's technology provides a digital readout of the amount of transcript in a sample. A pure digital readout of transcript counts is linear across a large dynamic range, exhibits less background noise and is less ambiguous for downstream analysis than technologies that use analog signals. Finally, the time, effort, and sample requirements of the nCounter system are more scalable than real-time PCR or microarrays. For example, to measure 500 genes using 2 ng of RNA per real-time PCR reaction in triplicate, one would need 3μg of total RNA and 1500 reactions whereas the same experiment could be performed using the nCounter system with 300ng of total RNA in 3 reactions.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1044

There are many applications for a technology that is capable of highly-multiplexed measurement of gene expression from relatively low amounts of starting material, particularly one which can detect transcripts of low abundance. For example, estimates of mRNA expression levels in both mouse and human cells suggest that the vast majority of the genes in the transcriptome are expressed at or below 20 transcripts per cell.[7, 8] Currently, real-time PCR is the most widely accepted platform for measuring low-abundance messages. We have shown the nCounter system yields remarkable similar results. Another potential application of the technology is to measure expression profiles in clinical settings. Several studies have used expression arrays to identify a set of genes whose expression pattern or "signature" can serve as a clinical diagnostic or prognostic indicator. Classic examples of such studies include the AML/ALL work of Golub et al.[9] and the breast cancer classification studies of van't Veer et al.[10, 11] After identifying a set of predictive genes via full genome arrays, one would like to validate their expression profile on a large number of patients and ultimately develop a diagnostic assay (see Simon[12] for a recent review). Typically these clinical signatures involve more than 30, but fewer than 500, genes. The nCounter system is ideally suited for profiling such clinically-relevant signatures, particularly from small samples with limited amounts of RNA such as tissue biopsies, micro-dissected or laser-captured samples, and cells sorted by flow cytometry. Preliminary work using the nCounter system directly on cell lysates looks promising, and has the potential to reduce further the amount of sample and sample handling needed.

REFERENCES CITED

1. Pruitt, K.D., Tatusova, T. & Maglott, D.R. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **33**, D501-504 (2005).
2. Shi, L. et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* **24**, 1151-1161 (2006).
3. Kuo, W.P. et al. A sequence-oriented comparison of gene expression measurements across different hybridization-based technologies. *Nat Biotechnol* **24**, 832-840 (2006).
4. Patterson, T.A. et al. Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. *Nat Biotechnol* **24**, 1140-1150 (2006).

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1045

5. Canales, R.D. et al. Evaluation of DNA microarray results with quantitative gene expression platforms. *Nat Biotechnol* **24**, 1115-1122 (2006).
6. Oliveri, P., Carrick, D.M. & Davidson, E.H. A regulatory gene network that directs micromere specification in the sea urchin embryo. *Dev Biol* **246**, 209-228 (2002).
7. Hastie, N.D. & Bishop, J.O. The expression of three abundance classes of messenger RNA in mouse tissues. *Cell* **9**, 761-774 (1976).
8. Velculescu, V.E. et al. Analysis of human transcriptomes. *Nat Genet* **23**, 387-388 (1999).
9. Golub, T.R. et al. Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. *Science* **286**, 531-537 (1999).
10. van 't Veer, L.J. et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530-536 (2002).
11. van de Vijver, M.J. et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* **347**, 1999-2009 (2002).
12. Simon, R. Roadmap for developing and validating therapeutically relevant genomic classifiers. *J Clin Oncol* **23**, 7332-7341 (2005).
13. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J Mol Biol* **215**, 403-410 (1990).
14. Kane, M.D. et al. Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res* **28**, 4552-4557 (2000).
15. Li, X., He, Z. & Zhou, J. Selection of optimal oligonucleotide probes for microarrays using multiple criteria, global alignment and parameter estimation. *Nucleic Acids Res* **33**, 6114-6123 (2005).
16. Gentleman, R.C. et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80 (2004).
17. Brazma, A. et al. ArrayExpress--a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res* **31**, 68-71 (2003).
18. Rast, J.P. et al. Recovery of developmentally defined gene sets from high-density cDNA macroarrays. *Dev Biol* **228**, 270-286 (2000).

The present invention can be implemented as a computer program product that comprises a computer program mechanism embedded in a computer readable storage medium. For instance, the computer program product could contain the program modules shown in Figure 19. These program modules can be stored on a CD-ROM, DVD, magnetic disk storage product, or any other computer readable data or program storage product. The program modules can also be embedded in permanent storage, such as ROM, one or more programmable chips, or one or more application specific integrated circuits (ASICs). Such permanent storage can be localized in a server, 802.11 access point, 802.11 wireless bridge/station, repeater, router, mobile phone, or other electronic devices. The program modules in the computer program product can also be distributed electronically, via the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1046

Internet or otherwise, by transmission of a computer data signal (in which the software modules are embedded) either digitally or on a carrier wave.

Many modifications and variations of this invention can be made without departing from its spirit and scope, as will be apparent to those skilled in the art. For instance, data storage module 44, label identification module 50, and probe identification module 54 can be combined into a single program, can each be a separate program, or could, in fact, be dispersed in multiple (*e.g.*, three or more) programs. The specific embodiments described herein are offered by way of example only, and the invention is to be limited only by the terms of the appended claims, along with the full scope of equivalents to which such claims are entitled.

All of the U.S. patents, U.S. patent application publications, U.S. patent applications, foreign patents, foreign patent applications and non-patent publications referred to in this specification and/or listed in the Application Data Sheet, are incorporated herein by reference, in their entirety. Aspects of the embodiments can be modified, if necessary to employ concepts of the various patents, applications and publications to provide yet further embodiments.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1047

WHAT IS CLAIMED IS:

1. A method for identifying a pair of adjacent target-specific sequences for use in a probe pair hybridizable to a target mRNA, comprising the steps of:

(a) generating a first pool of candidate nucleotide sequences of a first predetermined length or lengths that are reverse complements of a target mRNA sequence, wherein each candidate nucleotide sequence can be divided into two adjacent nucleotide sequences of equal length consisting of a 5′ candidate sequence and a 3′ candidate sequence;

(b) deleting from said first pool one or more candidate nucleotide sequences that meet at least two of the following criteria: (i) contain inverted repeats of greater than a predetermined length of consecutive nucleotides; (ii) contain direct repeats of greater than a predetermined length of consecutive nucleotides; (iii) whose 5′ candidate sequence and/or 3′ candidate sequence have a GC content outside a predetermined range; (iv) whose 5′ candidate sequence and/or 3′ candidate sequence contain contiguous stretches of C residues of greater than a predetermined length; and (v) whose 5′ candidate sequence and/or 3′ candidate sequence have melting temperatures that are outside a first predetermined melting temperature range; thereby generating a second pool of candidate nucleotide sequences;

(c) deleting from said second pool one or more candidate nucleotide sequences whose 5′ candidate sequence and/or 3′ candidate sequence has a cross-hybridization potential to non-specific sequences that is higher than a predetermined threshold, thereby generating a third pool of candidate nucleotide sequences;

(d) deleting from said third pool one or more candidate nucleotide sequences whose 5′ candidate sequence and/or 3′ candidate sequence has a melting temperature outside a second predetermined temperature range, wherein the second predetermined melting temperature range is within the first predetermined melting temperature range;

(e) determining the melting temperature for a modified 5′ candidate sequence or a modified 3′ candidate sequence, wherein the modified 5′ candidate sequence or a modified 3′ candidate sequence is a modified form of a 5′ candidate sequence or a 3′

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1048

candidate sequence, respectively, of a candidate nucleotide sequence deleted in step (d) because its 5′ candidate sequence and/or 3′ candidate sequence has a melting temperature above the second predetermined range, wherein the modified 5′ candidate sequence has been modified by trimming at least one nucleotide from the 5′ end of the corresponding 5′ candidate sequence, and wherein the modified 3′ candidate sequence has been modified by trimming at least one nucleotide from the 3′ end of the corresponding 3′ candidate sequence;

(f) in the event that:

(A) the modified 5′ or modified 3′ candidate sequence, and

(B) a 3′ or 5′, respectively, candidate sequence or the modified form thereof;

each have a melting temperature within the second predetermined melting temperature range and both are derived from the same candidate nucleotide sequence;

adding to the third pool a modified candidate nucleotide sequence composed of (A) and (B), thereby generating a fourth pool of candidate nucleotide sequences;

(g) in the event that the length of the modified 5′ or modified 3′ candidate sequence is greater than a second predetermined length, repeating step (e) one or more times wherein the modified 5′ candidate sequence or modified 3′ candidate sequence, respectively, has been trimmed by a greater number of nucleotides than in step (e) each time, until the length of the modified 5′ or modified 3′ candidate sequence is the earlier of (i) equal to, or (ii) lower than, the second predetermined length;

(h) for each modified 5′ or modified 3′ candidate sequence of step (g) wherein:

(C) said modified 5′ or modified 3′ candidate sequence, and

(D) a 3′ or 5′, respectively, candidate sequence or the modified form thereof;

each have a melting temperature within the second predetermined melting temperature range and both are derived from the same candidate nucleotide sequence;

adding to the third pool a modified candidate sequence composed of (C) and (D), thereby generating a fifth pool of candidate nucleotide sequences; and

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1049

(i) optionally repeating steps (e)-(h) for one or more different candidate nucleotide sequences deleted in step (d),

thereby generating a sixth pool of candidate nucleotide sequences,

whereby the fourth, fifth and sixth pools consist of candidate nucleotide sequences composed of pairs of adjacent target-specific sequences for use in a probe pair hybridizable to the target mRNA.

2. The method of claim 1, which further comprises the step of outputting to a user interface device, a computer readable storage medium, or a local or remote computer system, or displaying, one or a plurality of candidate nucleotide sequences and/or modified candidate nucleotide sequences in the fourth, fifth and/or sixth pools and/or the 5′ candidate sequences or modified 5′ candidate sequences and/or 3′ candidate sequences or modified 3′ candidate sequences contained therein.

3. The method of claim 2, wherein the candidate nucleotide sequences and/or modified candidate nucleotide sequences are outputted as pairs of adjacent target-specific nucleotide sequences derived from said candidate nucleotide sequences and/or modified candidate nucleotide sequences, respectively.

4. The method of any one of claims 1 to 3, wherein in step (c) a 5′ candidate sequence or 3′ candidate sequence is deemed to have a cross-hybridization potential to non-specific sequences that is higher than said predetermined threshold if said 5′ candidate sequence or 3′ candidate sequence has (i) a sequence percentage identity with a first sequence (hereinafter "first non-target sequence") or its complement that is equal to or greater than a first predetermined cutoff, said first non-target sequence being other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA, and said first non-target sequence being present in a database comprising cellular mRNA sequences or cDNA sequences derived therefrom; and (ii) a contiguous block of sequence identity with a second sequence (hereinafter "second non-target sequence") or its complement that is equal

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1050

to or greater than a second predetermined cutoff, said second non-target sequence being other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA, and said second non-target sequence being present in the database.

5. The method of any one of claims 1 to 4, wherein a plurality of candidate nucleotide sequences and/or modified candidate nucleotide sequences in the fourth, fifth and/or sixth pools and/or the 5′ candidate sequences or modified 5′ candidate sequences and 3′ candidate sequences or modified 3′ candidate sequences contained therein are outputted or displayed in a ranked order based on a weighted score of the cross-hybridization potentials and the melting temperatures of said 5′ candidate sequences or modified 5′ candidate sequences and 3′ candidate sequences or modified 3′ candidate sequences.

6. The method of claim 5, wherein the weighted score is calculated according to the formula:

(Tm score * WFa) + (MCB score * WFb) + (PID score * WFc)

where:

Tm score is a melting temperature score calculated according to the formula:

(differential score + general score)/3

where the differential score is calculated according to the following formula:

1-|(TmA - TmB)| / (TmHco - TmLco)

where the general score is calculated according to the following formula:

(((TmI - |(TmA - TmI)|) / TmI) + (((TmI - |(TmB - TmI)|) / TmI)))

where TmA is the melting temperature of the 5′ candidate sequence or modified 5′ candidate sequence in a pair of adjacent target-specific sequences, TmB is the melting temperature of the 3′ candidate sequence or modified 3′ candidate sequence in said pair of adjacent target-specific sequences, TmHco is the upper limit of the second

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1051

predetermined temperature range; TmLco is the lower limit of the second predetermined temperature range; and TmI is a predetermined ideal melting temperature;

where:

MCB score is a maximum contiguous block score calculated according to the formula:

$$1 - (MCB / MCBco);$$

where MCB is the greater of:

(i) the maximum contiguous block of identity between:

(A) a first target-specific nucleotide sequence in said pair of adjacent target-specific sequences; and

(B) a sequence in the database other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA;

and

(ii) the maximum contiguous block of identity between:

(A) a second target-specific nucleotide sequence in said pair of adjacent target-specific sequences; and

(B) a sequence in the database other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA,

and wherein MCBco is the first predetermined cutoff;

where:

PID score is a percent identity score calculated according to the formula:

$$1 - (PID / PIDco));$$

where PID is the greater of:

(i) the greatest percentage sequence identity between:

(A) a first target-specific nucleotide sequence in said pair of adjacent target-specific sequences; and

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1052

(B) a sequence in the database other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA;

and

(ii) the greatest percentage sequence identity between:

(A) a second target-specific nucleotide sequence in said pair of adjacent target-specific sequences; and

(B) a sequence in the database other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA,

and wherein PIDco is the second predetermined cutoff,

and where WFa, WFb, and WFc are each independently a weighting factor, each of which is a real number.

7. A method for identifying a pair of adjacent target-specific sequences for use in a probe pair hybridizable to a target mRNA, comprising the steps of:

(a) generating a first pool of candidate nucleotide sequences of 100 nucleotides that are reverse complements of a target mRNA sequence, wherein each candidate nucleotide sequence can be divided into two adjacent nucleotide sequences of 50 nucleotides each, said adjacent nucleotide sequences consisting of a 5′ candidate sequence and a 3′ candidate sequence;

(b) deleting from said first pool one or more candidate nucleotide sequences that meet the following criteria: (i) contain inverted repeats that are 6 consecutive nucleotides in length or greater; (ii) contain direct repeats that are 9 consecutive nucleotides in length or greater; (iii) whose 5′ candidate sequence and/or 3′ candidate sequence have a GC content outside 40-70%; (iv) whose 5′ candidate sequence and/or 3′ candidate sequence contain contiguous stretches of 3 C residues or greater; and (v) whose 5′ candidate sequence and/or 3′ candidate sequence have melting temperatures that are outside a range of (A) 60-90°C or (B) 65-85°C; thereby generating a second pool of candidate nucleotide sequences;

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1053

(c) deleting from said second pool one or more candidate nucleotide sequences whose 5′ candidate sequence and/or 3′ candidate sequence has (i) a sequence percentage identity of 85% or greater with a first sequence (hereinafter "first non-target sequence") or its complement, said first non-target sequence being other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA, and said first non-target sequence being present in a database comprising cellular mRNA sequences or cDNA sequences derived therefrom; and (ii) a contiguous block of sequence identity of 15 nucleotides or greater with a second sequence (hereinafter "second non-target sequence") or its complement, said second non-target sequence being other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA, and said second non-target sequence being present in the database; thereby generating a third pool of candidate nucleotide sequences;

(d) deleting from said third pool one or more candidate nucleotide sequences whose 5′ candidate sequence and/or 3′ candidate sequence has a melting temperature outside the range of 78-83°C;

(e) determining the melting temperature for a modified 5′ candidate sequence or a modified 3′ candidate sequence, wherein the modified 5′ candidate sequence or a modified 3′ candidate sequence is a modified form of a 5′ candidate sequence or a 3′ candidate sequence, respectively, of a candidate nucleotide sequence deleted in step (d) because its 5′ candidate sequence and/or 3′ candidate sequence has a melting temperature above 83°C, wherein the modified 5′ candidate sequence has been modified by trimming at least one nucleotide from the 5′ end of the corresponding 5′ candidate sequence, and wherein the modified 3′ candidate sequence has been modified by trimming at least one nucleotide from the 3′ end of the corresponding 3′ candidate sequence;

(f) in the event that:

(A) the modified 5′ or modified 3′ candidate sequence, and

(B) a 3′ or 5′, respectively, candidate sequence or the modified form thereof,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1054

each have a melting temperature within the range of 78-83°C and both are derived from the same candidate nucleotide sequence,
adding to the third pool a modified candidate nucleotide sequence composed of (A) and (B); thereby generating a fourth pool of candidate nucleotide sequences;

(g) in the event that the length of the modified 5′ or modified 3′ candidate sequence is greater than 35 nucleotides, repeating step (e) one or more times wherein the modified 5′ candidate sequence or modified 3′ candidate sequence, respectively, has been trimmed by a greater number of nucleotides than in step (e) each time, until the length of the modified 5′ or modified 3′ candidate sequence is the earlier of (i) equal to, or (ii) lower than, 35 nucleotides;

(h) for each modified 5′ or modified 3′ candidate sequence of step (g) wherein:

(C) the modified 5′ or modified 3′ candidate sequence, and

(D) a 3′ or 5′, respectively, candidate sequence or modified candidate sequence;

each have a melting temperature in the range of 78-83°C and both are derived from the same candidate nucleotide sequence,
adding to the third pool a modified candidate sequence composed of (C) and (D); thereby generating a fifth pool of candidate nucleotide sequences; and

(i) optionally repeating steps (e)-(h) for one or more different candidate nucleotide sequences deleted in step (d),
thereby generating a sixth pool of candidate nucleotide sequences,
whereby the fourth, fifth and sixth pools consist of candidate nucleotide sequences composed of pairs of adjacent target-specific sequences for use in a probe pair hybridizable to the target mRNA.

8. The method of claim 7, which further comprises the step of outputting to a user interface device, a computer readable storage medium, or a local or remote computer system, or displaying, one or a plurality of candidate nucleotide sequences and/or modified candidate nucleotide sequences in the fourth, fifth and/or sixth pools and/or the 5′

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1055

candidate sequences or modified 5′ candidate sequences and/or 3′ candidate sequences or modified 3′ candidate sequences contained therein.

9. The method of claim 8, wherein the candidate nucleotide sequences and/or modified candidate nucleotide sequences are outputted as a pair of adjacent target-specific nucleotide sequences derived from said candidate nucleotide sequences and/or modified candidate nucleotide sequences, respectively.

10. The method of any one of claims claim 7 to 9, wherein a plurality of candidate nucleotide sequences and/or modified candidate nucleotide sequences in the fourth, fifth and/or sixth pools and/or the 5′ candidate sequences or modified 5′ candidate sequences and 3′ candidate sequences or modified 3′ candidate sequences contained therein are outputted or displayed in a ranked order based on a weighted score of the cross-hybridization potentials and the melting temperatures of said 5′ candidate sequences or modified 5′ candidate sequences and 3′ candidate sequences or modified 3′ candidate sequences.

11. The method of claim 10, wherein the weighted score is calculated according to the formula:

(Tm score * WFa) + (MCB score * WFb) + (PID score * WFc)

where:

Tm score is a melting temperature score calculated according to the formula:

(differential score + general score)/3

where the differential score is calculated according to the following formula:

1-|(TmA - TmB)| / 5

where the general score is calculated according to the following formula:

(((80.5 - |(TmA - 80.5)|) / 80.5) + (((80.5 - |(TmB - 80.5)|) / 80.5)))

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1056

where TmA is the melting temperature of the 5′ candidate sequence or modified 5′ candidate sequence of a pair of adjacent target-specific sequences and TmB is the melting temperature of the 3′ candidate sequence or modified 3′ candidate sequence of said pair of adjacent target-specific sequences;

where:

MCB score is a maximum contiguous block score calculated according to the formula:

$$1 - (MCB / 15);$$

where MCB is the greater of:

(i) the maximum contiguous block of identity between:

(A) a first target-specific nucleotide sequence in said pair of adjacent target-specific sequences; and

(B) a sequence in the database other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA;

and

(ii) the maximum contiguous block of identity between:

(A) a second target-specific nucleotide sequence in said pair of adjacent target-specific sequences; and

(B) a sequence in the database other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA;

where:

PID score is a percent identity score calculated according to the formula:

$$1 - (PID / 85\%);$$

where PID is the greater of:

(i) the greatest percentage sequence identity between:

(A) a first target-specific nucleotide sequence in said pair of adjacent target-specific sequences; and

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1057

(B) a sequence in the database other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA;

and

(ii) the greatest percentage sequence identity between:

(A) a second target-specific nucleotide sequence in said pair of adjacent target-specific sequences; and

(B) a sequence in the database other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA,

and wherein PIDco is the second predetermined cutoff,

and where WFa, WFb, and WFc are each independently a weighting factor, each of which is a real number.

12. A method for identifying a plurality of pairs of adjacent target-specific sequences for use in a respective plurality of probe pairs, each probe pair being hybridizable to a different target mRNA, comprising, for each target mRNA: identifying a pair of adjacent target-specific sequences according to the method of any one of claims 1 to 11.

13. A method for identifying a target-specific nucleotide sequence for use in a probe hybridizable to a target mRNA, comprising the steps of:

(a) generating a first pool of candidate nucleotide sequences of a first predetermined length or lengths that are reverse complements of a target mRNA sequence;

(b) deleting from said first pool one or more candidate nucleotide sequences that meet at least two of the following criteria: (i) contains inverted repeats of greater than a predetermined length of consecutive nucleotides; (ii) contains direct repeats of greater than a predetermined length of consecutive nucleotides; (iii) has a GC content outside a predetermined range; (iv) contains a contiguous stretch of C residues of greater than a predetermined length; and (v) has a melting temperature that is outside a first predetermined

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1058

melting temperature range; thereby generating a second pool of candidate nucleotide sequences;

(c) deleting from said second pool one or more candidate nucleotide sequences that have a cross-hybridization potential to non-specific sequences that is higher than a predetermined threshold, thereby generating a third pool of candidate n sequences;

(d) deleting from said third pool one or more candidate nucleotide sequences that have a melting temperature outside a second predetermined temperature range, wherein the second predetermined melting temperature range is within the first predetermined melting temperature range;

(e) determining the melting temperature for a modified candidate nucleotide sequence, wherein the modified candidate nucleotide sequence is a modified form of a candidate nucleotide sequence deleted in step (d) because it has a melting temperature above the second predetermined range, wherein the modified candidate nucleotide sequence has been modified by trimming at least one nucleotide from the 5′ end or the 3′ end of said candidate nucleotide sequence;

(f) in the event that the modified candidate nucleotide sequence has a melting temperature within the second predetermined melting temperature range, adding to the third pool the modified candidate nucleotide sequence, thereby generating a fourth pool of candidate nucleotide sequences;

(g) in the event that the length of the modified candidate nucleotide sequence is greater than a second predetermined length, repeating step (e) one or more times wherein the modified candidate nucleotide sequence has been trimmed by a greater number of nucleotides than in step (e) each time, until the length of the modified candidate nucleotide sequence is the earlier of (i) equal to, or (ii) lower than, the second predetermined length;

(h) adding to the third pool each modified candidate nucleotide sequence of step (g) which has a melting temperature within the second predetermined melting temperature range; thereby generating a fifth pool of candidate nucleotide sequences; and

(i) optionally repeating steps (e)-(h) for one or more different candidate nucleotide sequences deleted in step (d), thereby generating a sixth pool of candidate nucleotide sequences,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1059

whereby the fourth, fifth and sixth pools consist of target-specific nucleotide sequences for use in a probe hybridizable to a target mRNA.

14. The method of claim 13, which further comprises the step of outputting to a user interface device, a computer readable storage medium, or a local or remote computer system, or displaying, one or a plurality of candidate nucleotide sequences and/or modified candidate nucleotide sequences in the fourth, fifth and/or sixth pools.

15. The method of any one of claims 13 and 14, wherein in step (c) a candidate target-specific sequence is deemed to have a cross-hybridization potential to non-specific sequences that is higher than said predetermined threshold if said candidate target-specific sequence has (i) a sequence percentage identity with a first sequence (hereinafter "first non-target sequence") or its complement that is equal to or greater than a first predetermined cutoff, said first non-target sequence being other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA, and said first non-target sequence being present in a database comprising cellular mRNA sequences or cDNA sequences derived therefrom; and (ii) a contiguous block of sequence identity with a second sequence (hereinafter "second non-target sequence") or its complement that is equal to or greater than a second predetermined cutoff, said second non-target sequence being other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA, and said second non-target sequence being present in the database.

16. The method of any one of claims 13 and 14, wherein a plurality of candidate nucleotide sequences and/or modified candidate nucleotide sequences in the fourth, fifth and/or sixth pools are outputted or displayed in a ranked order based on a weighted score of the cross-hybridization potentials and the melting temperatures of said candidate nucleotide sequences and/or modified candidate nucleotide sequences.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1060

17. The method of claim 15, wherein a plurality of candidate nucleotide sequences and/or modified candidate nucleotide sequences in the fourth, fifth and/or sixth pools are outputted or displayed in a ranked order based on a weighted score of the cross-hybridization potentials and the melting temperatures of said candidate nucleotide sequences and/or modified candidate nucleotide sequences.

18. A method for identifying a plurality of target-specific sequences for use in a respective plurality of probes, each probe being hybridizable to a different target mRNA, comprising, for each target mRNA: identifying a target-specific sequence according to the method of any one of claims 1 to 17.

19. The method of any one of claims 1 to 18 wherein if the fourth, fifth and/or sixth pools contains no candidate nucleotide sequences, the method further comprises repeating steps (b) to (i), wherein step (b) is performed under more relaxed criteria.

20. The method of any one of claims 1 to 12, wherein in each step (g) of repeating the modified 5′ candidate sequence or modified 3′ candidate sequence is trimmed by one nucleotide.

21. The method of any one of claims 13 to 18, wherein in each step (g) of repeating the modified candidate nucleotide sequence is trimmed by one nucleotide.

22. The method of any one of claims 1 to 6 and 13 to 18, wherein step (b) comprises deleting from said first pool one or more candidate nucleotide sequences that meet at least 3 of said following criteria.

23. The method of claim 22, wherein step (b) comprises deleting from said first pool one or more candidate nucleotide sequences that meet 4 of said following criteria.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1061

24. The method of claim 22, wherein step (b) comprises deleting from said first pool one or more candidate nucleotide sequences that meet all 5 of said following criteria.

25. The method of claim any one of claims 1 to 6, wherein the first predetermined length is selected from the range of 70 to 120 nucleotides.

26. The method of claim 25, wherein the second predetermined length is selected from the range of 30 to 45 nucleotides.

27. The method of claim any one of claims 13 to 18, wherein the first predetermined length is selected from the range of 35 to 60 nucleotides.

28. The method of claim 27, wherein the second predetermined length is selected from the range of 30 to 45 nucleotides.

29. The method of any one of claims 1 to 6 and 13 to 18, wherein the predetermined length of the inverted repeats of step (b)(i) is selected from the range of 5 to 7 consecutive nucleotides.

30. The method of any one of claims 1 to 6 and 13 to 18, wherein the predetermined length of the direct repeats of step (b)(ii) is selected from the range of 7 to 9 consecutive nucleotides.

31. The method of any one of claims 1 to 6 and 13 to 18, wherein the predetermined range of GC content of step (b)(iii) is from 35-45% at the lower limit to 65-80% at the upper limit.

32. The method of claim 31, wherein the predetermined range of GC content of step (b)(iii) is 40-70%.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1062

33. The method of any one of claims 1 to 6 and 13 to 18, wherein the predetermined length in step (b)(iv) is 3.

34. The method of any one of claims 1 to 6 and 13 to 18, wherein the highest and lowest temperatures of the first predetermined melting temperature range differ by 20°C to 25°C.

35. The method of claim 34, wherein the first predetermined melting temperature range is from 60°C to 90°C.

36. The method of claim 34, wherein the first predetermined melting temperature range is from 65°C to 85°C.

37. The method of any one of claims 4, 6, 15, and 17, wherein the first predetermined cutoff is selected from the range of 70-95% sequence identity.

38. The method of claim 37, wherein the first predetermined cutoff is selected from the range of 80-90% sequence identity.

39. The method of any one of claims 4, 6, 15, 17, 37 and 38, wherein the second predetermined cutoff is selected from the range of 10-18 contiguous nucleotides.

40. The method of claim 39, wherein the second predetermined cutoff is selected from the range of 14-16 contiguous nucleotides.

41. The method of any one of claims 1 to 6 and 13 to 18, wherein the highest and lowest temperatures of the second predetermined melting temperature range differ by 4°C to 8°C.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1063

42. The method of claim 41, wherein the second predetermined melting temperature range is from 78°C to 83°C.

43. The method of any one of claims 4, 7, and 15, wherein the first non-target sequence and the second non-target sequence are the same.

44. The method of any one of claims 1 to 43, which further comprises the step of deleting from the fourth, fifth and/or sixth pools candidate nucleotide sequences that have a cross-hybridization potential to sequences present in other components of the probe or in a preparation step for the probe.

45. The method of any one of claims 1 to 43, wherein the target mRNA is an alternatively spliced mRNA.

46. The method of claim 45, which further comprises the step of determining whether one or more candidate nucleotide sequences are unique to one splice form or common to more than one splice form of the target mRNA.

47. The method of claim 45, wherein the first pool of candidate nucleotide sequences contains only candidate nucleotide sequences unique to one splice form.

48. The method of claim 45, wherein the first pool of candidate target-specific sequences contains only candidate nucleotide sequences common to multiple splice forms.

49. The method of any one of claims 1 to 48, which is a computer-implemented method.

50. A computer system for identifying a pair of adjacent target-specific sequences for use in a probe pair hybridizable to a target mRNA, comprising: a processor

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1064

and a memory coupled with the processor, the memory storing a plurality of machine instructions that cause the processor to perform the method of any one of claims 1 to 10.

51. A computer system for identifying a plurality of pairs of adjacent target-specific sequences for use in a respective plurality of probe pairs, each probe pair being hybridizable to a different target mRNA, comprising: a processor and a memory coupled with the processor, the memory storing a plurality of machine instructions that cause the processor to perform the method of claim 12.

52. A computer system for identifying a target-specific sequence for use in a probe hybridizable to a target mRNA, comprising: a processor and a memory coupled with the processor, the memory storing a plurality of machine instructions that cause the processor to perform the method of any one of claims 13 to 17.

53. A computer system for identifying a plurality of target-specific sequences for use in a respective plurality of probes, each probe being hybridizable to a different target mRNA, comprising: a processor and a memory coupled with the processor, the memory storing a plurality of machine instructions that cause the processor to perform the method of claim 18.

54. A computer system comprising: a processor and a memory coupled with the processor, the memory storing a plurality of machine instructions that cause the processor to perform the method of any one of claims 1 to 49.

55. A computer program product for use in conjunction with a computer system, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising instructions for performing the method of any one of claims 1 to 49.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1065

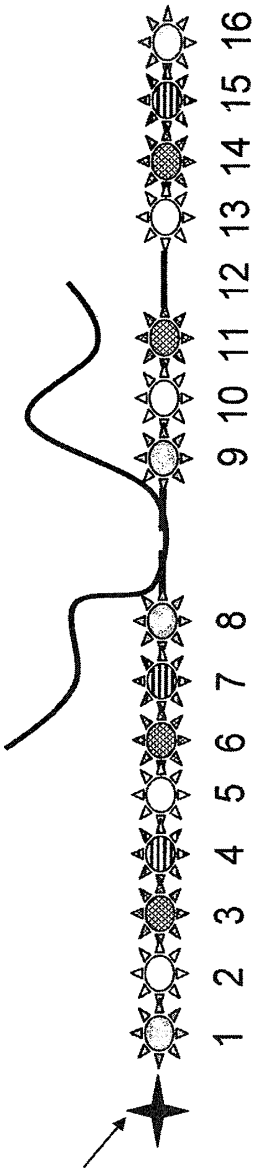


Figure 1A

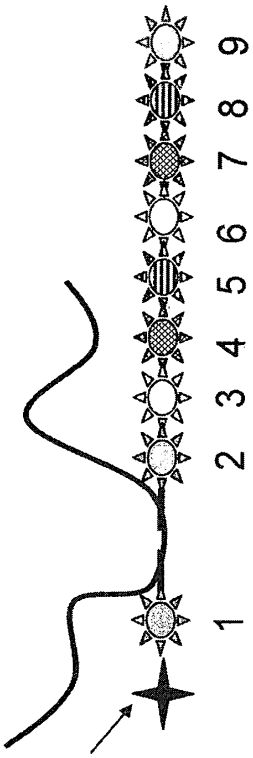


Figure 1B

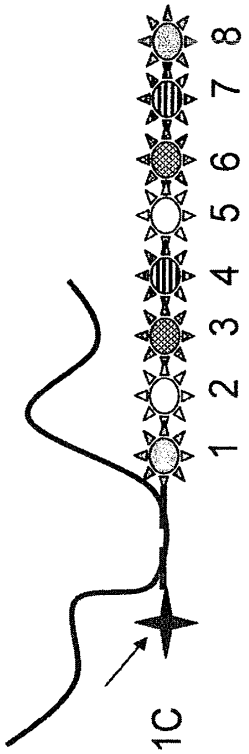


Figure 1C

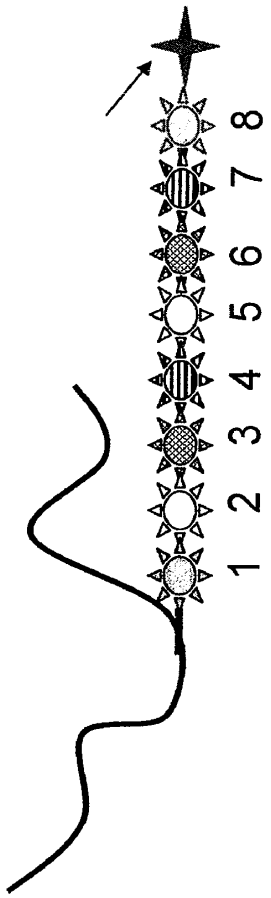


Figure 1D

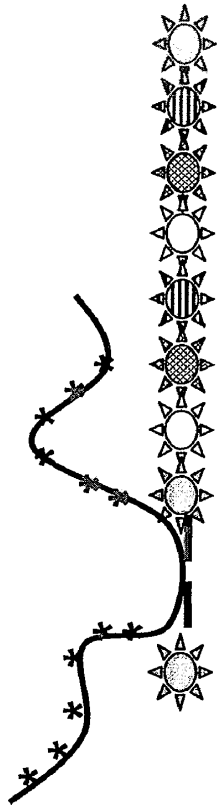


Figure 1E

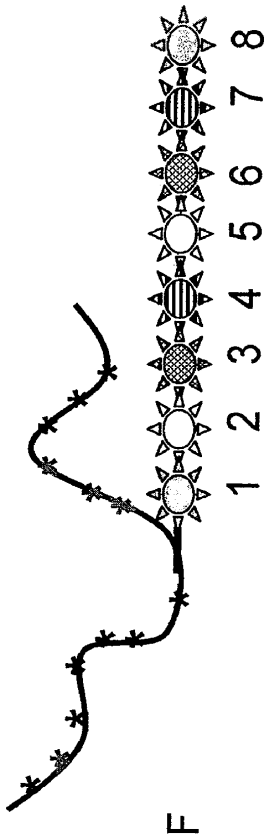


Figure 1F

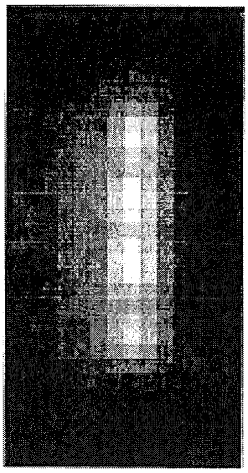


Figure 2C

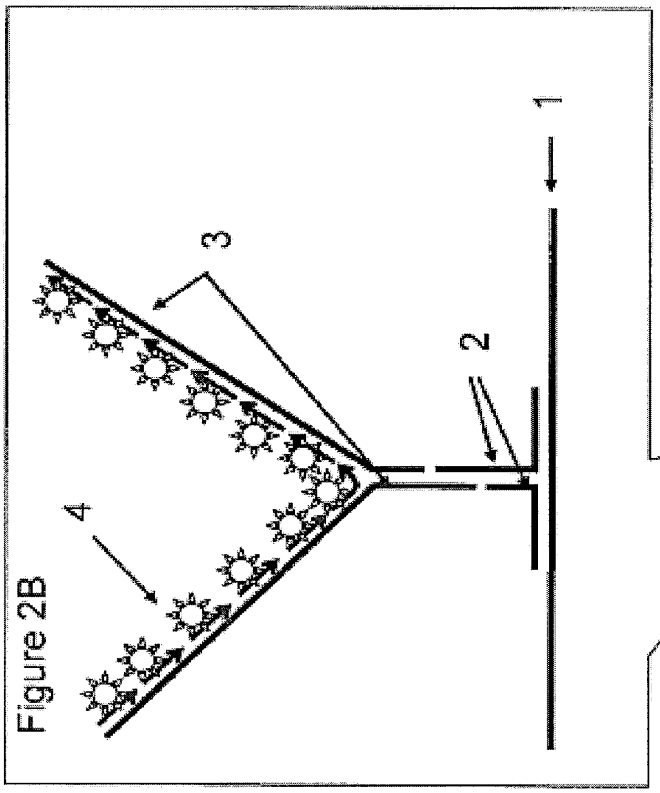


Figure 2B

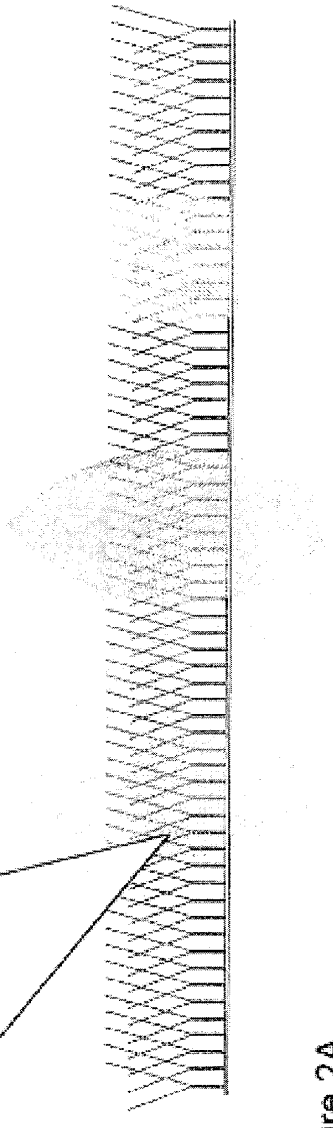


Figure 2A

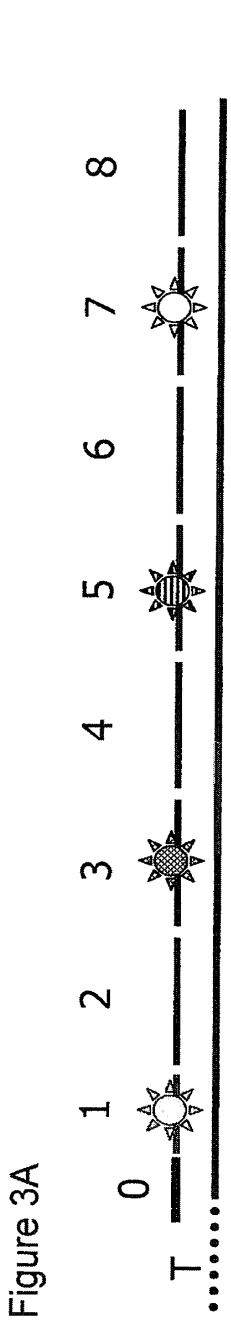


Figure 3A

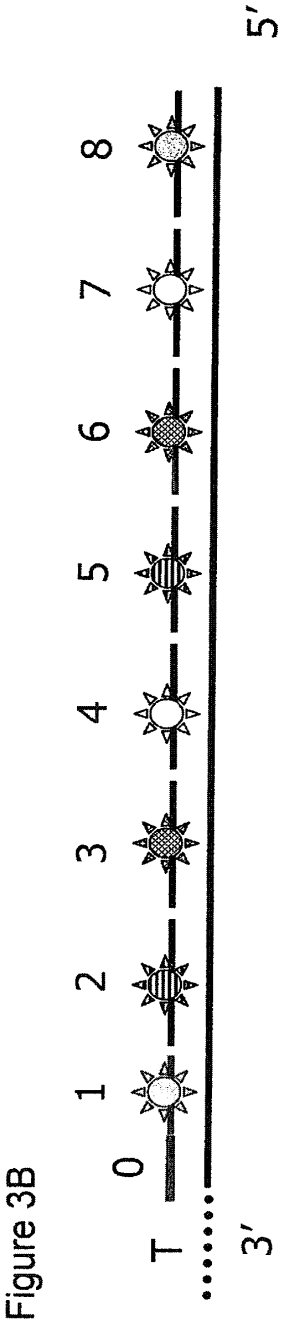


Figure 3B

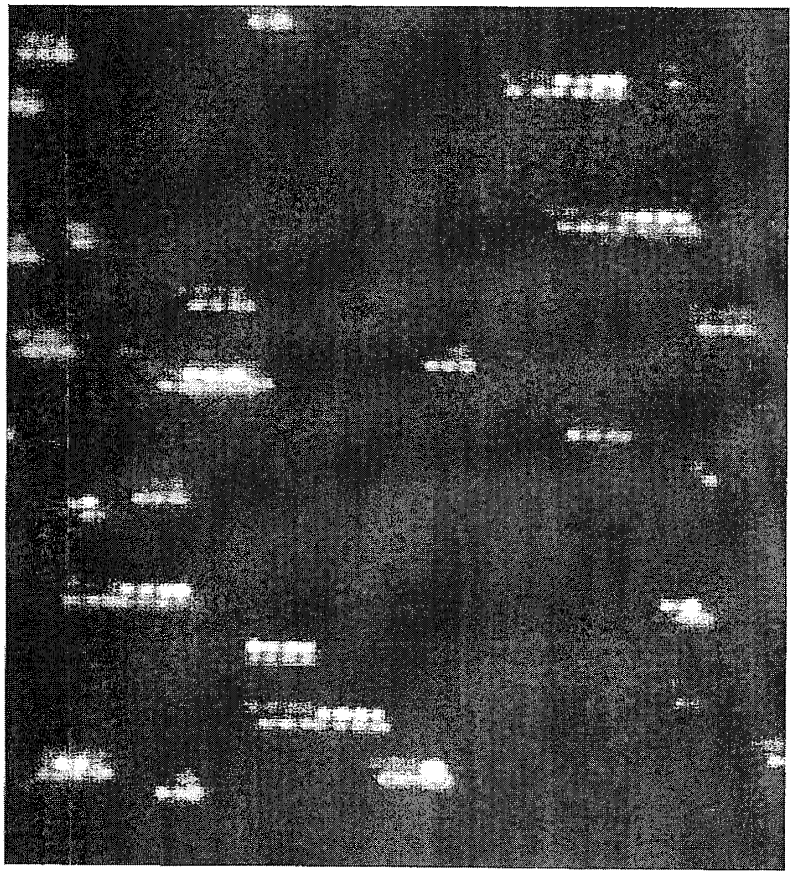
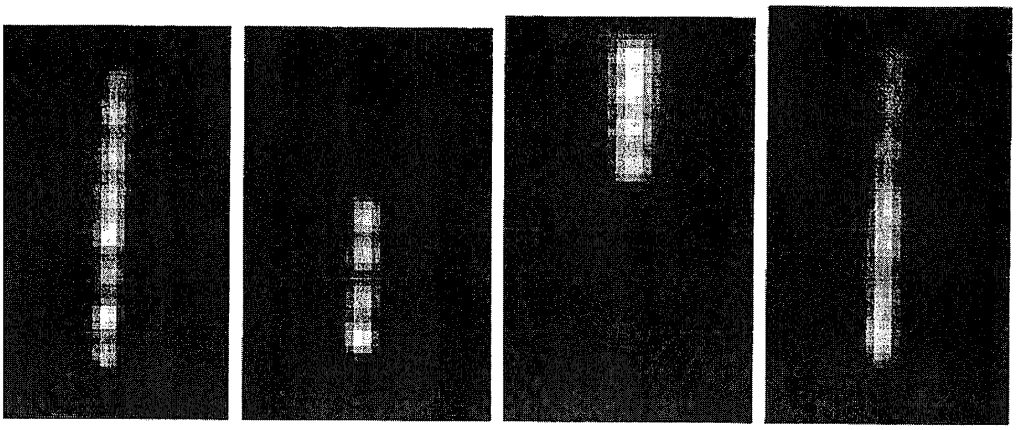


FIGURE 4



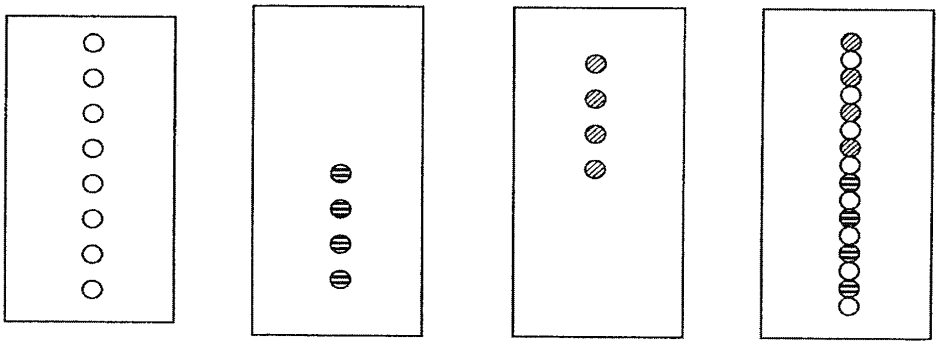


Figure 5A Figure 5B Figure 5C Figure 5D

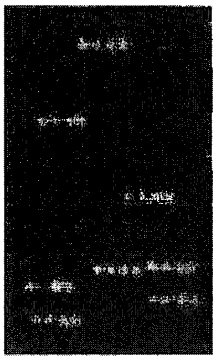


Figure 6E

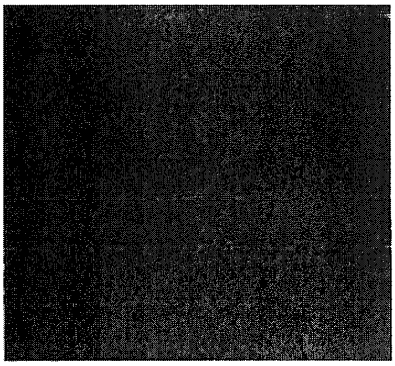


Figure 6D

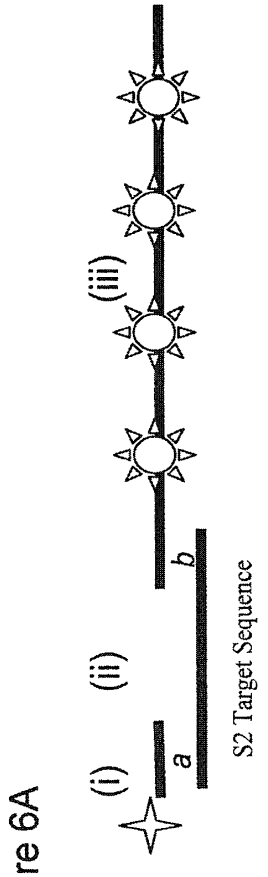


Figure 6A

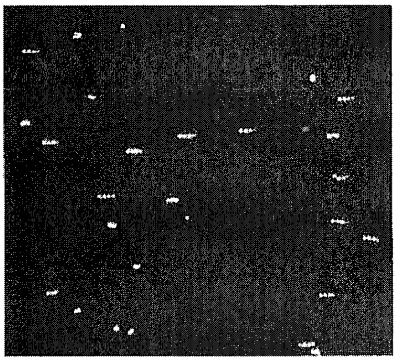


Figure 6C

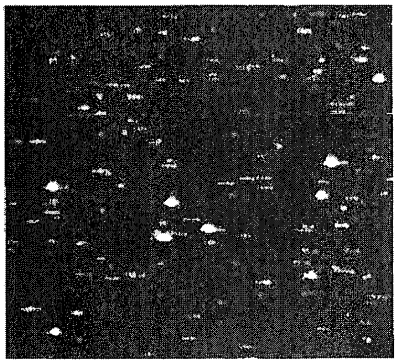


Figure 6B

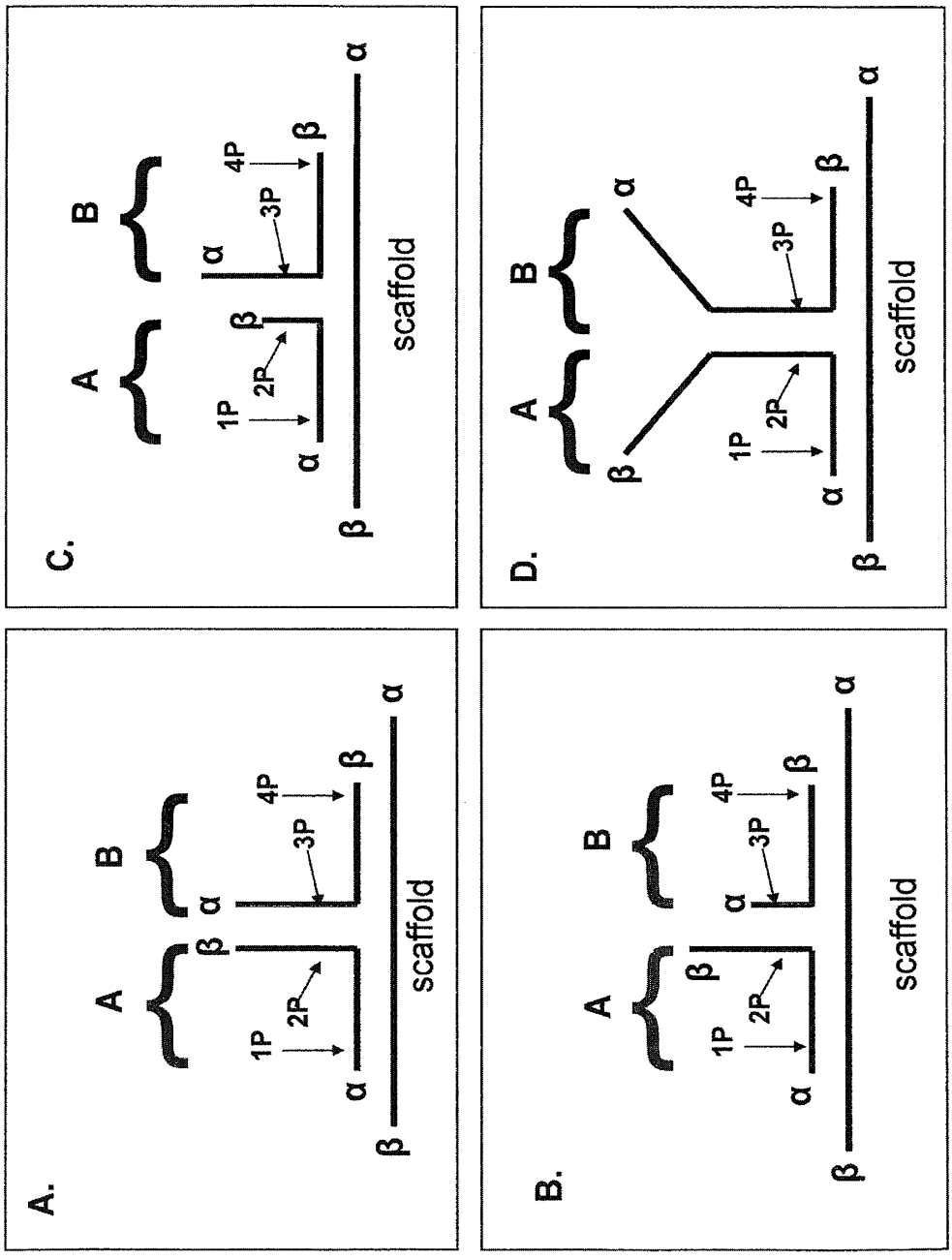


FIGURE 7A-D

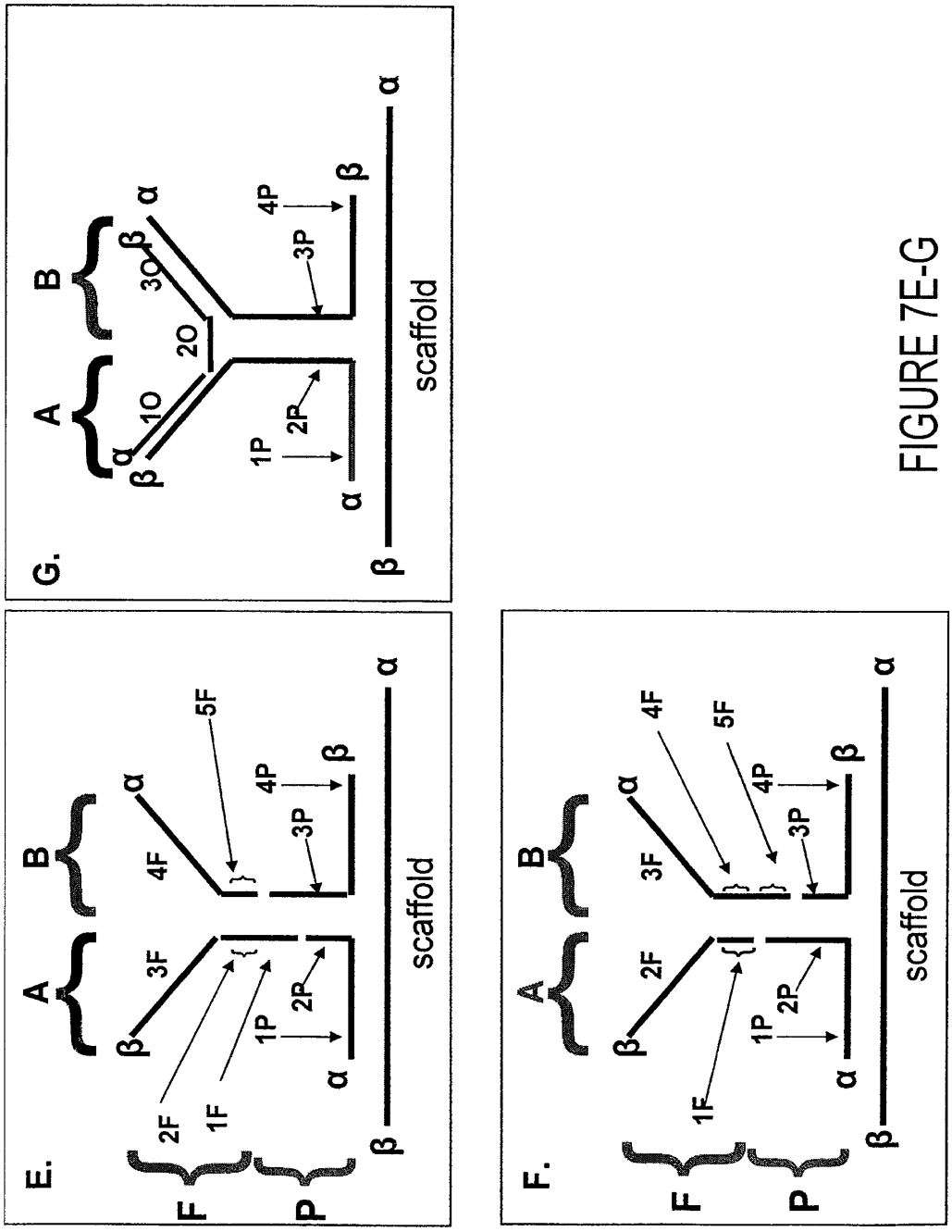


FIGURE 7E-G

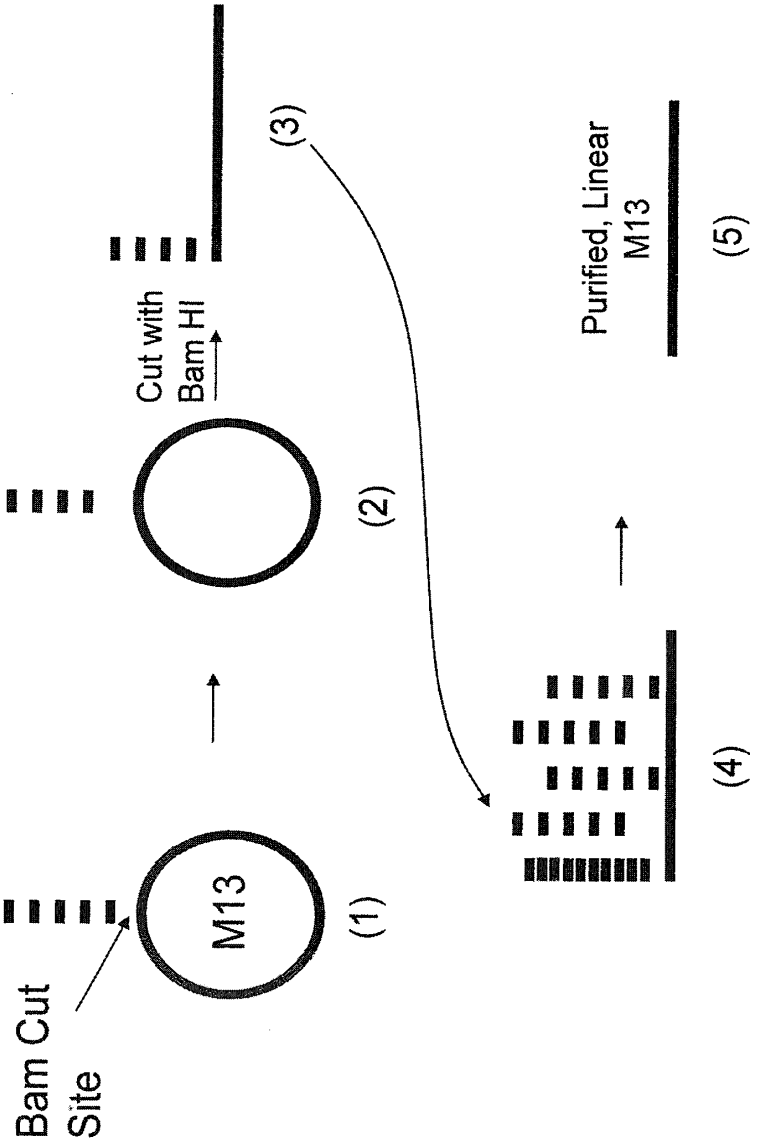


FIGURE 8

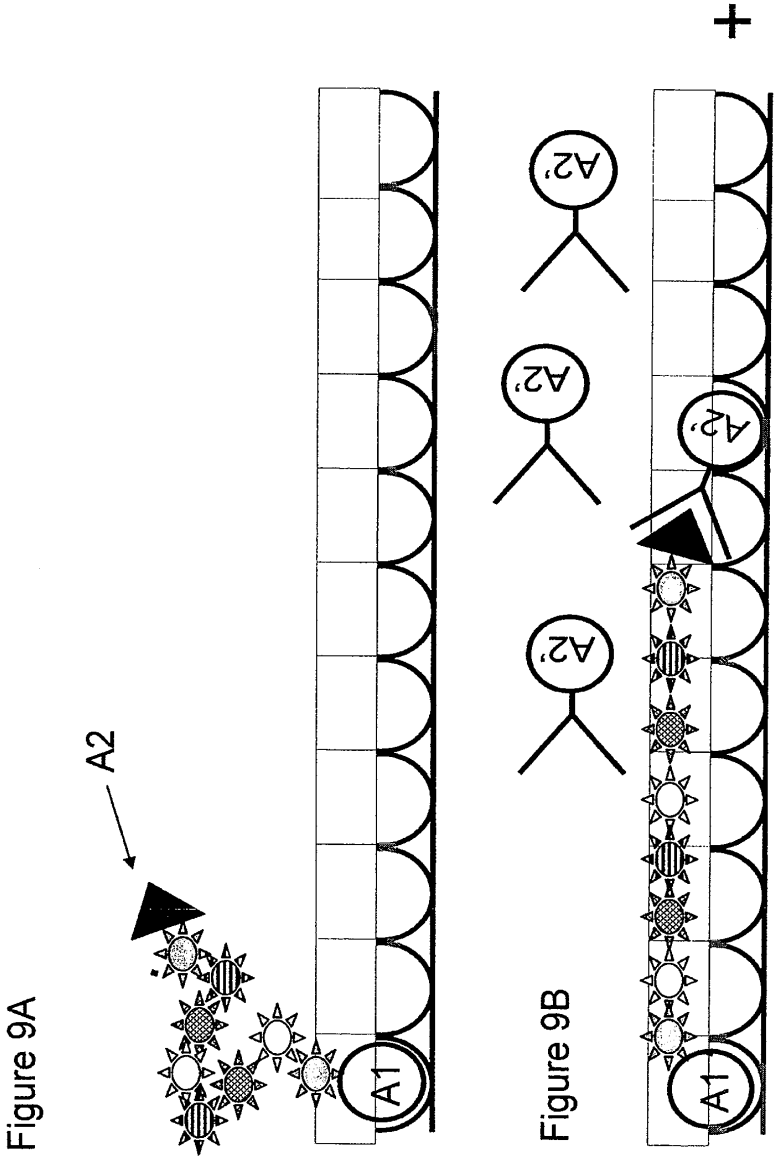


Figure 9A

Figure 9B

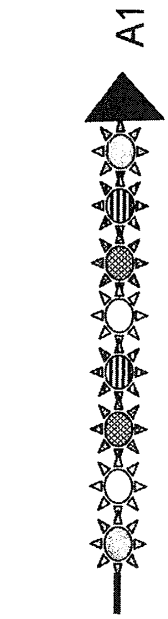


Figure 10A

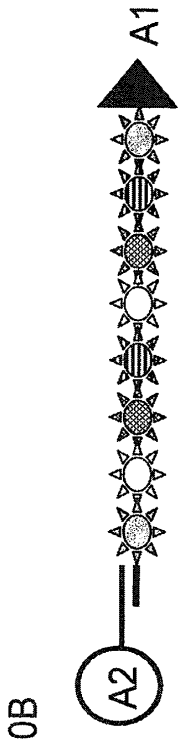


Figure 10B

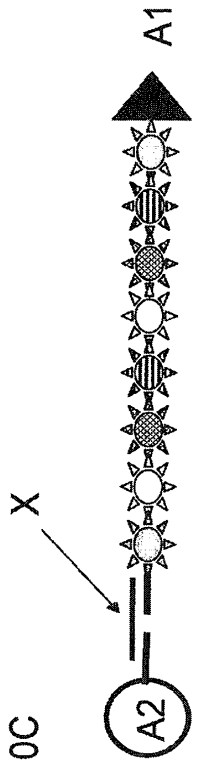


Figure 10C

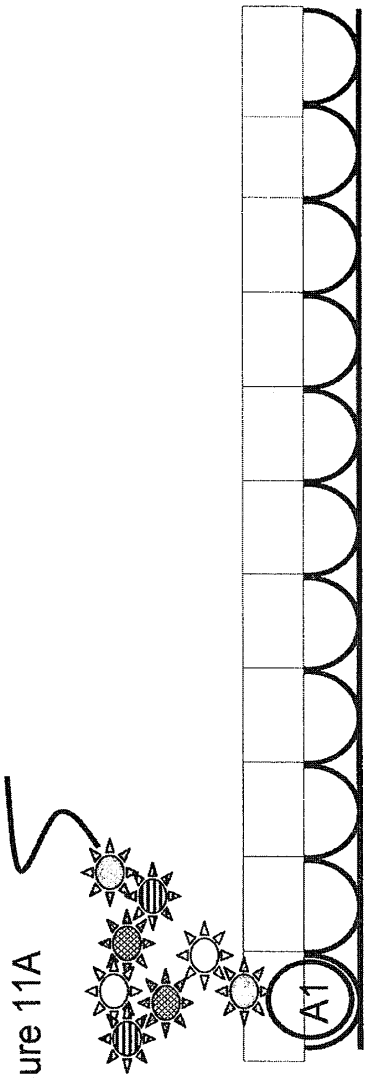


Figure 11A

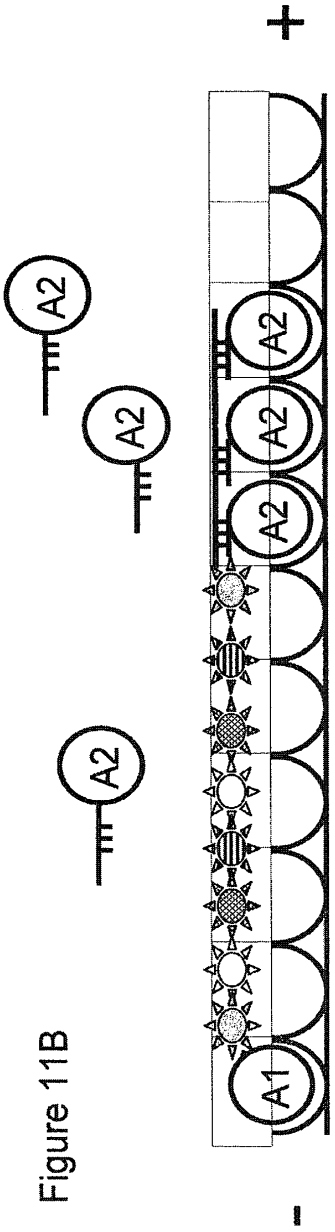


Figure 11B

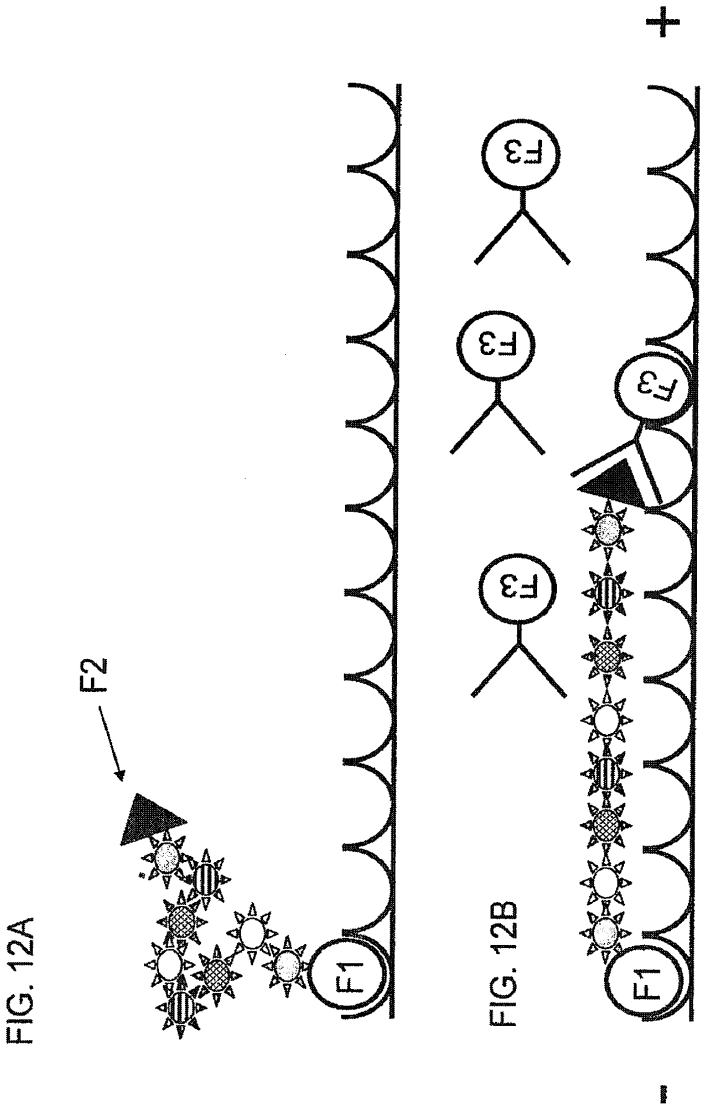


FIG. 12A

FIG. 12B

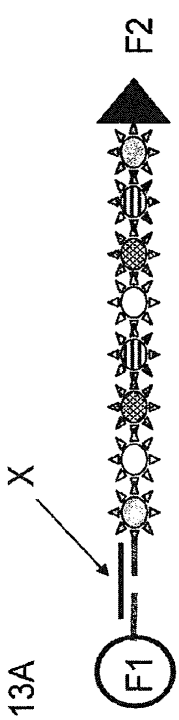


FIG. 13A

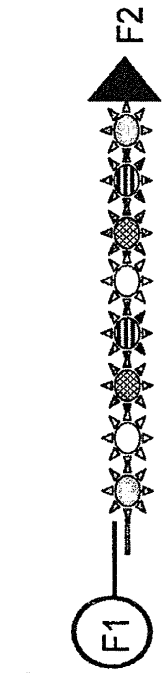


FIG. 13B



FIG. 13C

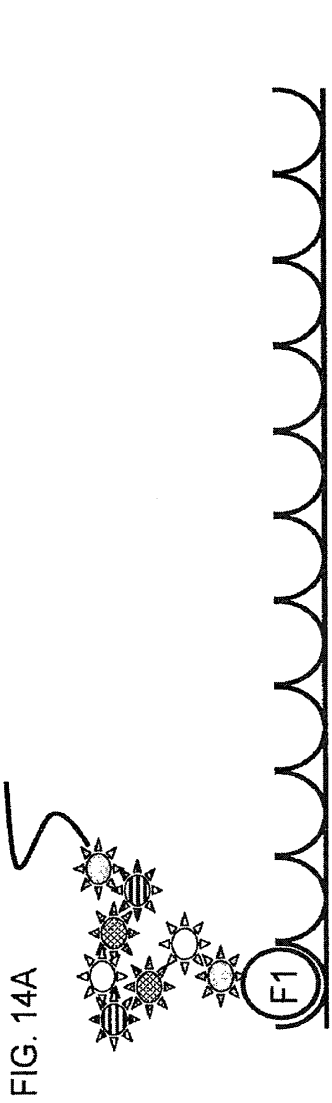


FIG. 14A

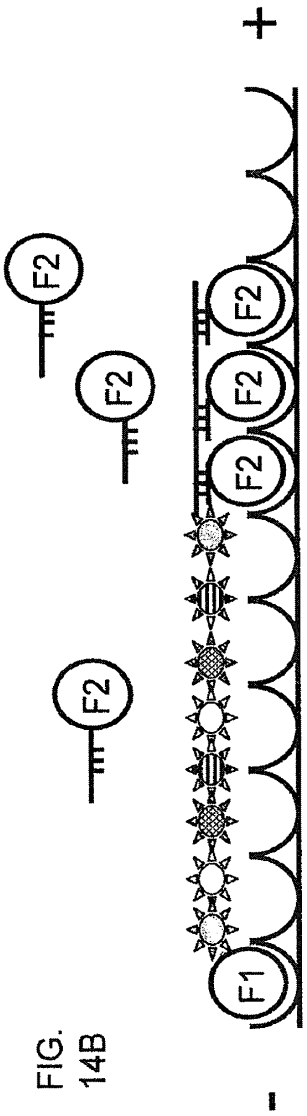


FIG. 14B

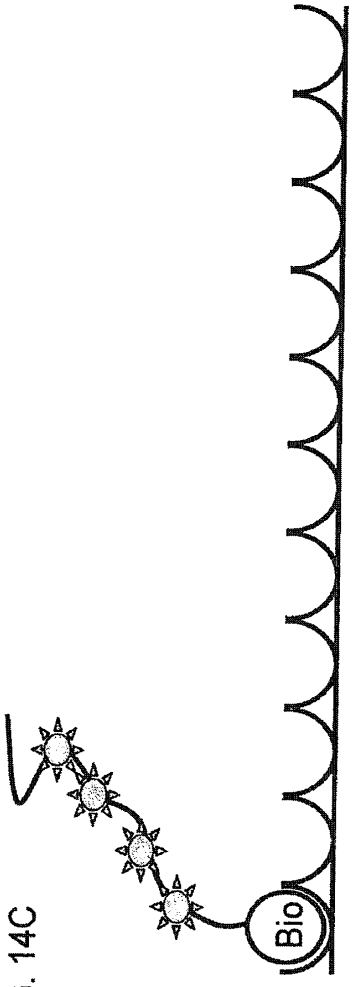


FIG. 14C

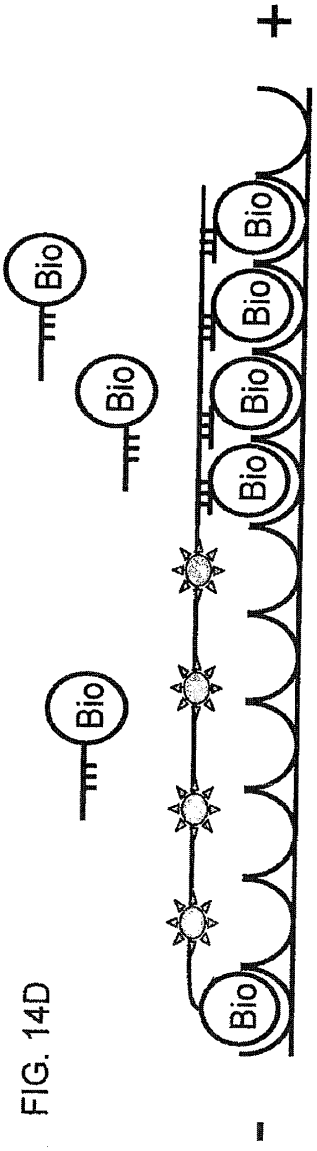


FIG. 14D

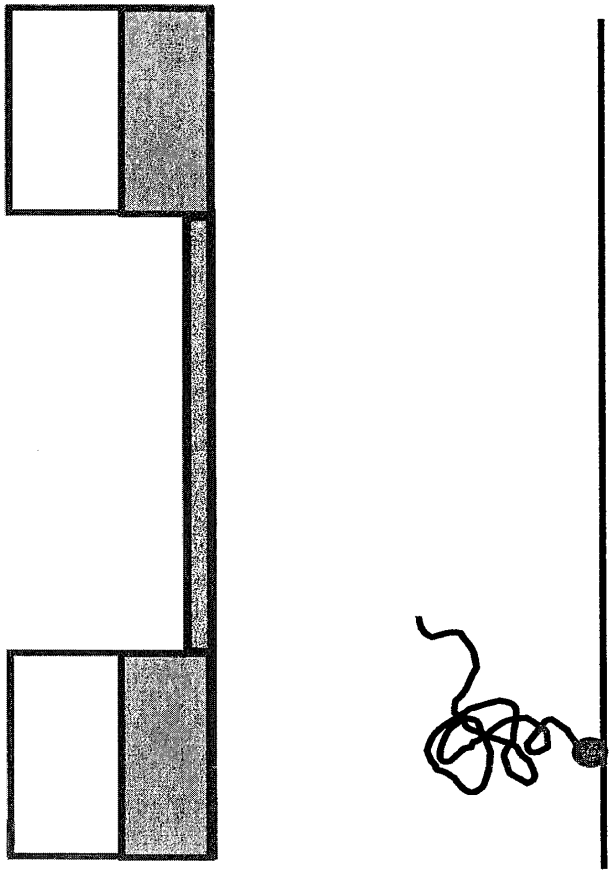


FIGURE 15A

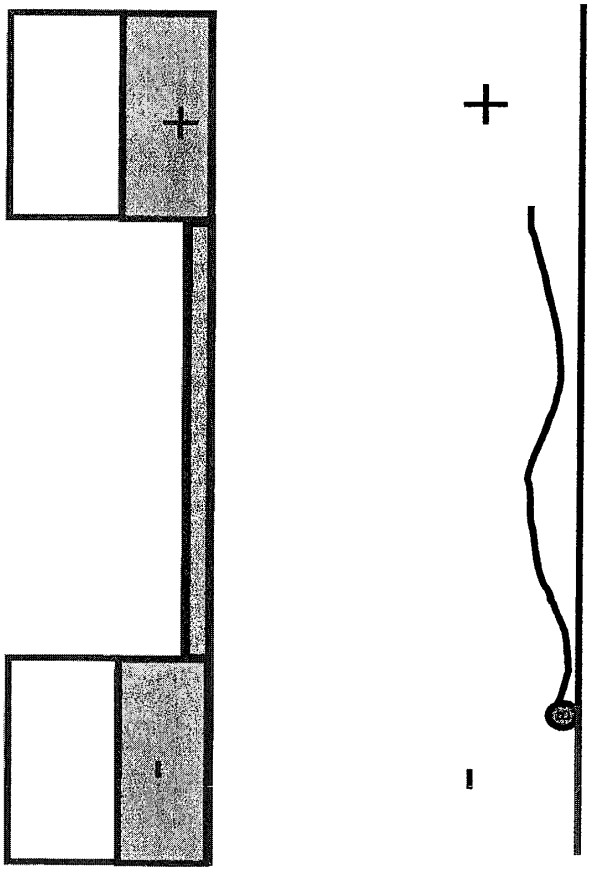


FIGURE 15B

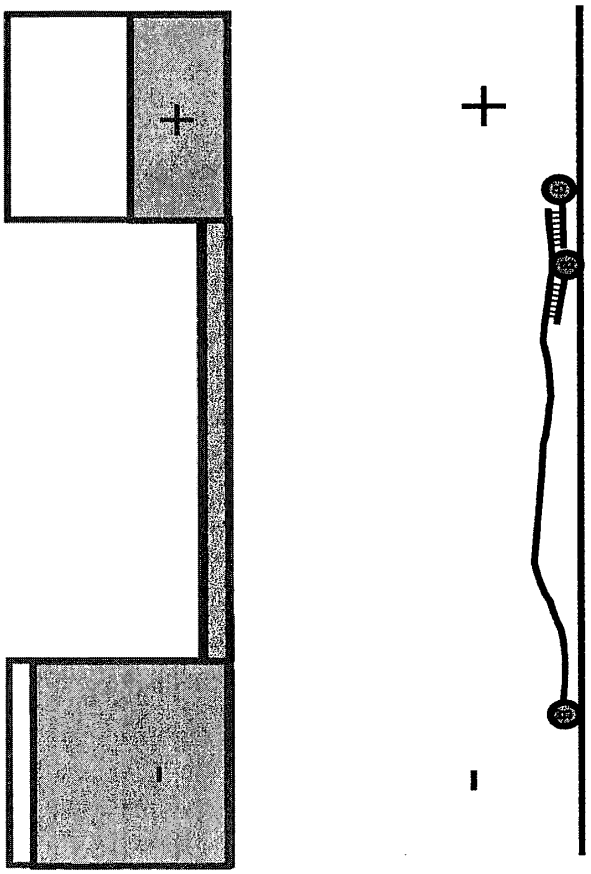


FIGURE 15C

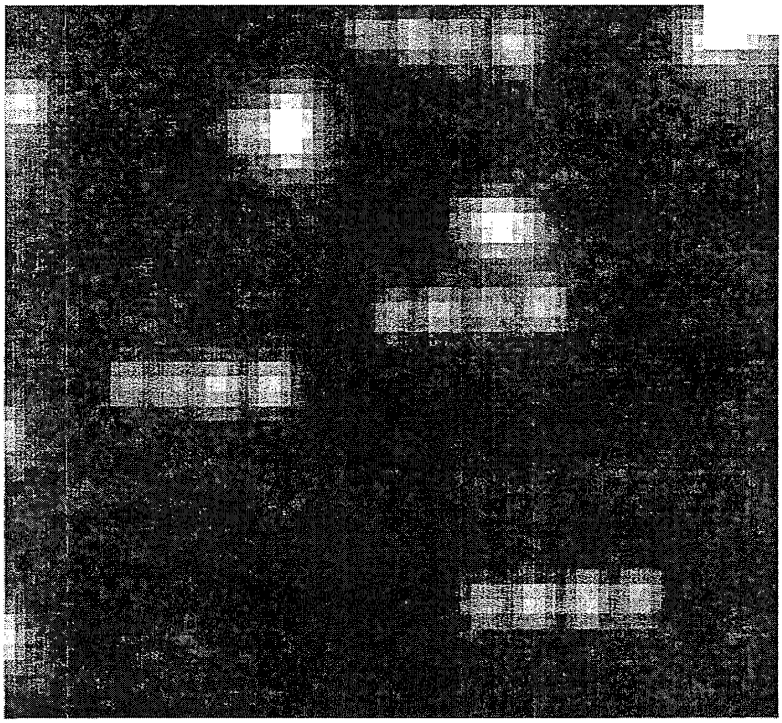


FIGURE 16

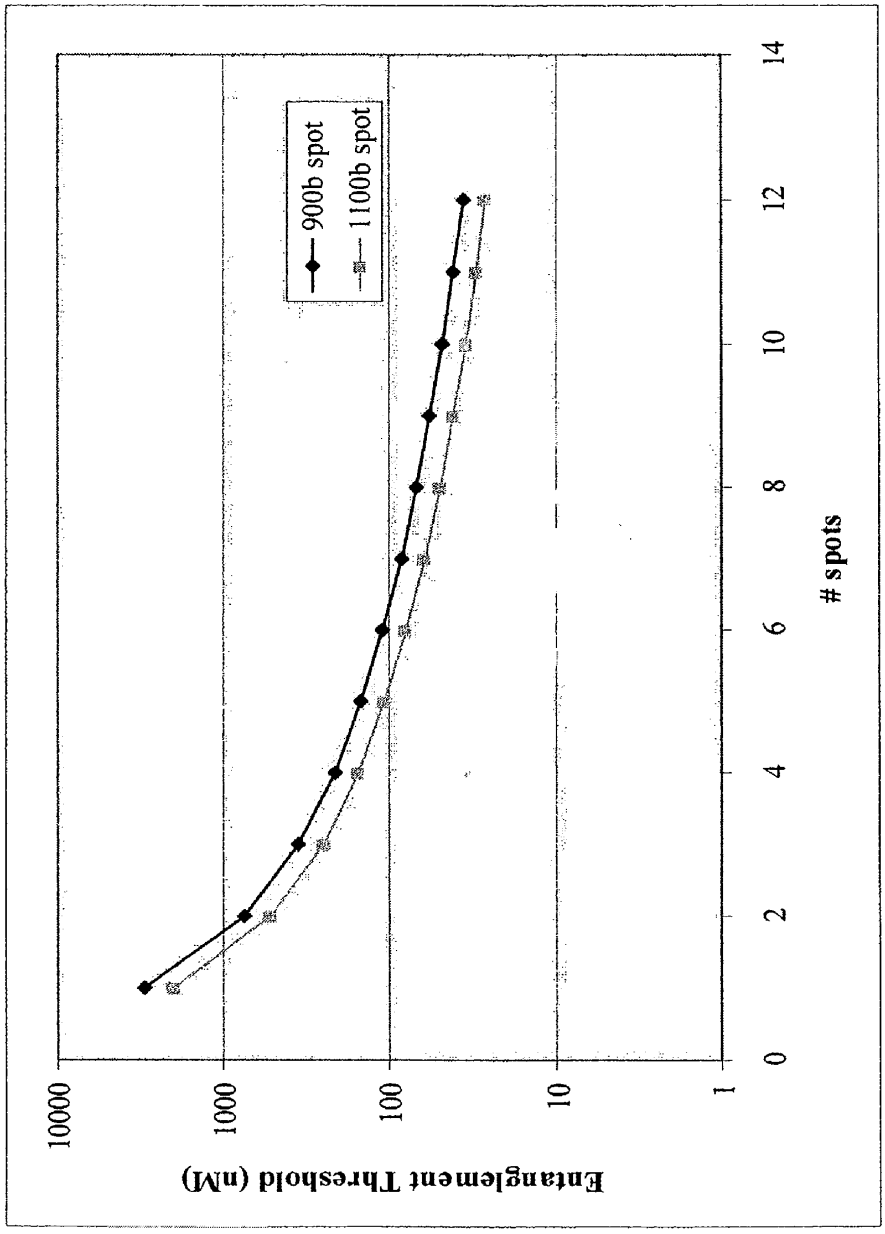

900b spot
1100b spot
Entanglement Threshold (nM)
10000
1000
100
10
1
0
2
4
6
8
10
12
14
spots

FIGURE 17

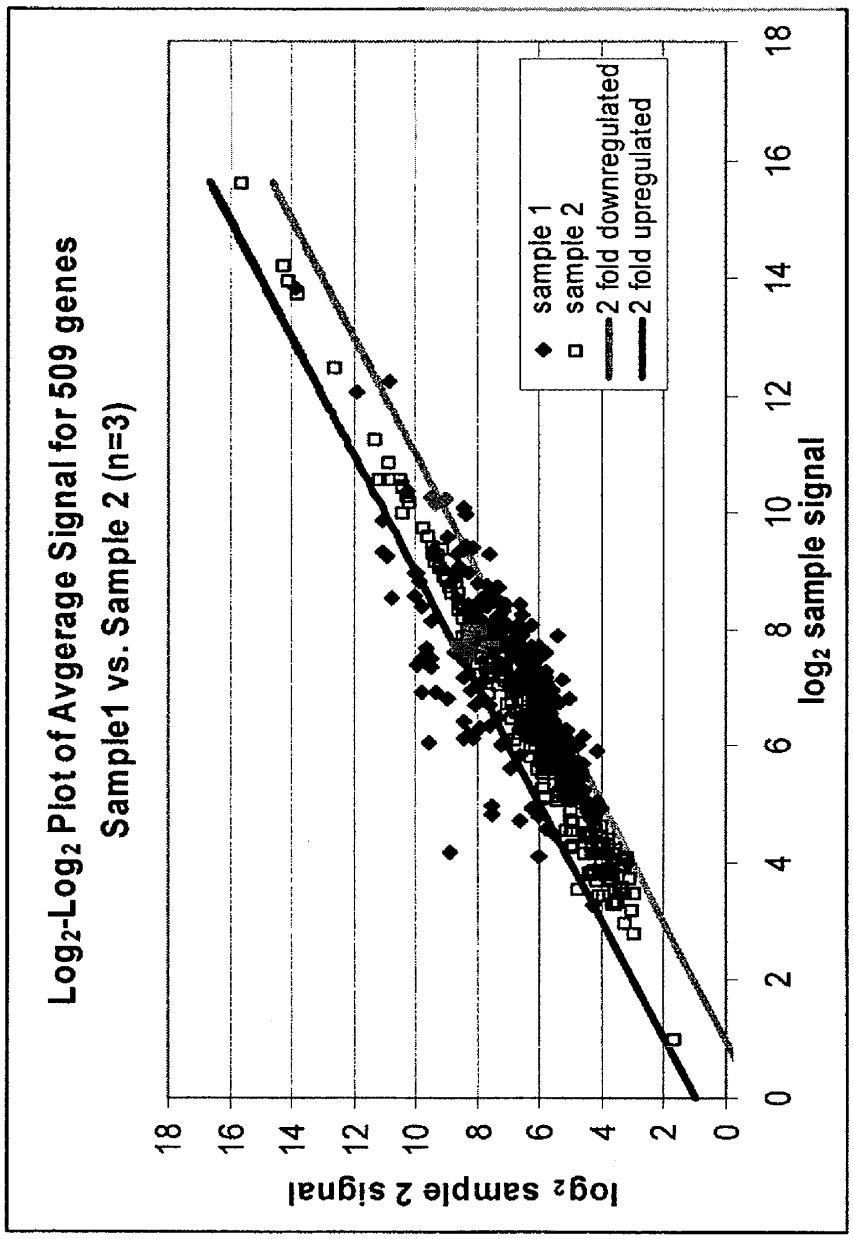

Log2-Log2 Plot of Avgerage Signal for 509 genes
Sample1 vs. Sample 2 (n=3)
log2 sample 2 signal
log2 sample signal
sample 1
sample 2
2 fold downregulated
2 fold upregulated
0
2
4
6
8
10
12
14
16
18

FIGURE 18

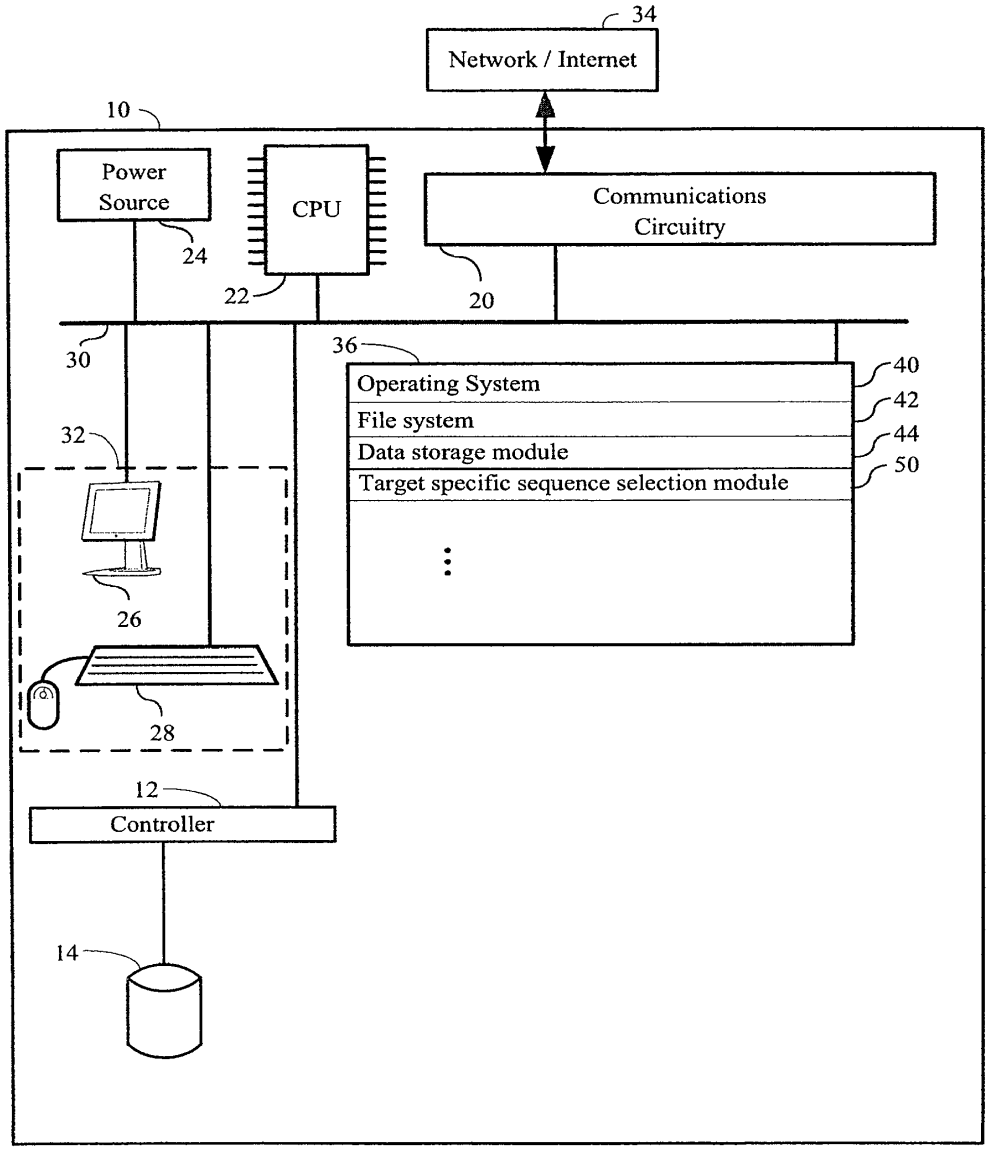

34
Network / Internet
10
Power Source
24
CPU
22
Communications Circuitry
20
30
36
Operating System
File system
Data storage module
Target specific sequence selection module
40
42
44
50
32
26
28
12
Controller
14

FIGURE 19

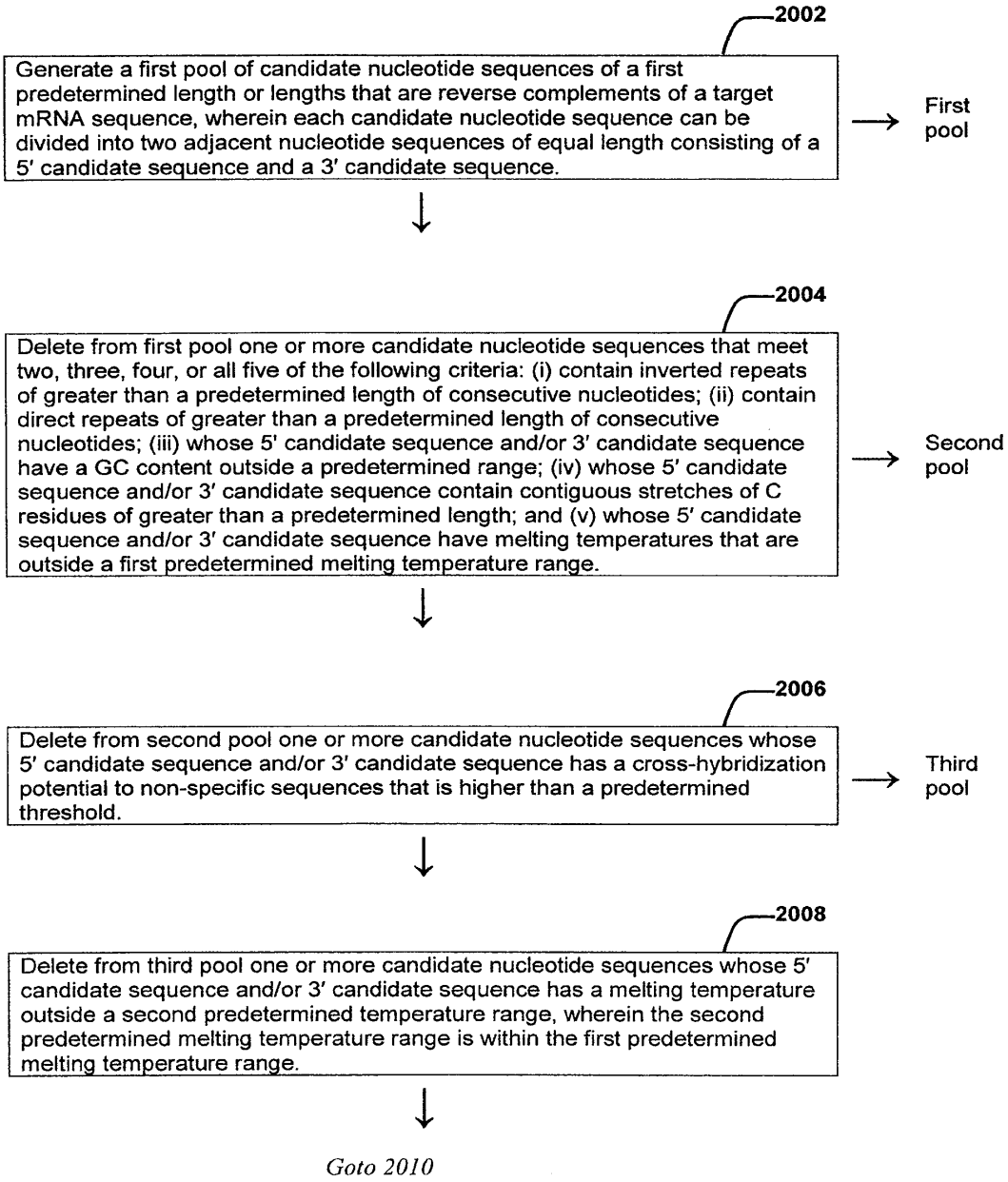

2002
Generate a first pool of candidate nucleotide sequences of a first predetermined length or lengths that are reverse complements of a target mRNA sequence, wherein each candidate nucleotide sequence can be divided into two adjacent nucleotide sequences of equal length consisting of a 5′ candidate sequence and a 3′ candidate sequence.
First pool
2004
Delete from first pool one or more candidate nucleotide sequences that meet two, three, four, or all five of the following criteria: (i) contain inverted repeats of greater than a predetermined length of consecutive nucleotides; (ii) contain direct repeats of greater than a predetermined length of consecutive nucleotides; (iii) whose 5' candidate sequence and/or 3′ candidate sequence have a GC content outside a predetermined range; (iv) whose 5′ candidate sequence and/or 3′ candidate sequence contain contiguous stretches of C residues of greater than a predetermined length; and (v) whose 5′ candidate sequence and/or 3′ candidate sequence have melting temperatures that are outside a first predetermined melting temperature range.
Second pool
2006
Delete from second pool one or more candidate nucleotide sequences whose 5′ candidate sequence and/or 3′ candidate sequence has a cross-hybridization potential to non-specific sequences that is higher than a predetermined threshold.
Third pool
2008
Delete from third pool one or more candidate nucleotide sequences whose 5′ candidate sequence and/or 3′ candidate sequence has a melting temperature outside a second predetermined temperature range, wherein the second predetermined melting temperature range is within the first predetermined melting temperature range.
Goto 2010

FIGURE 20A

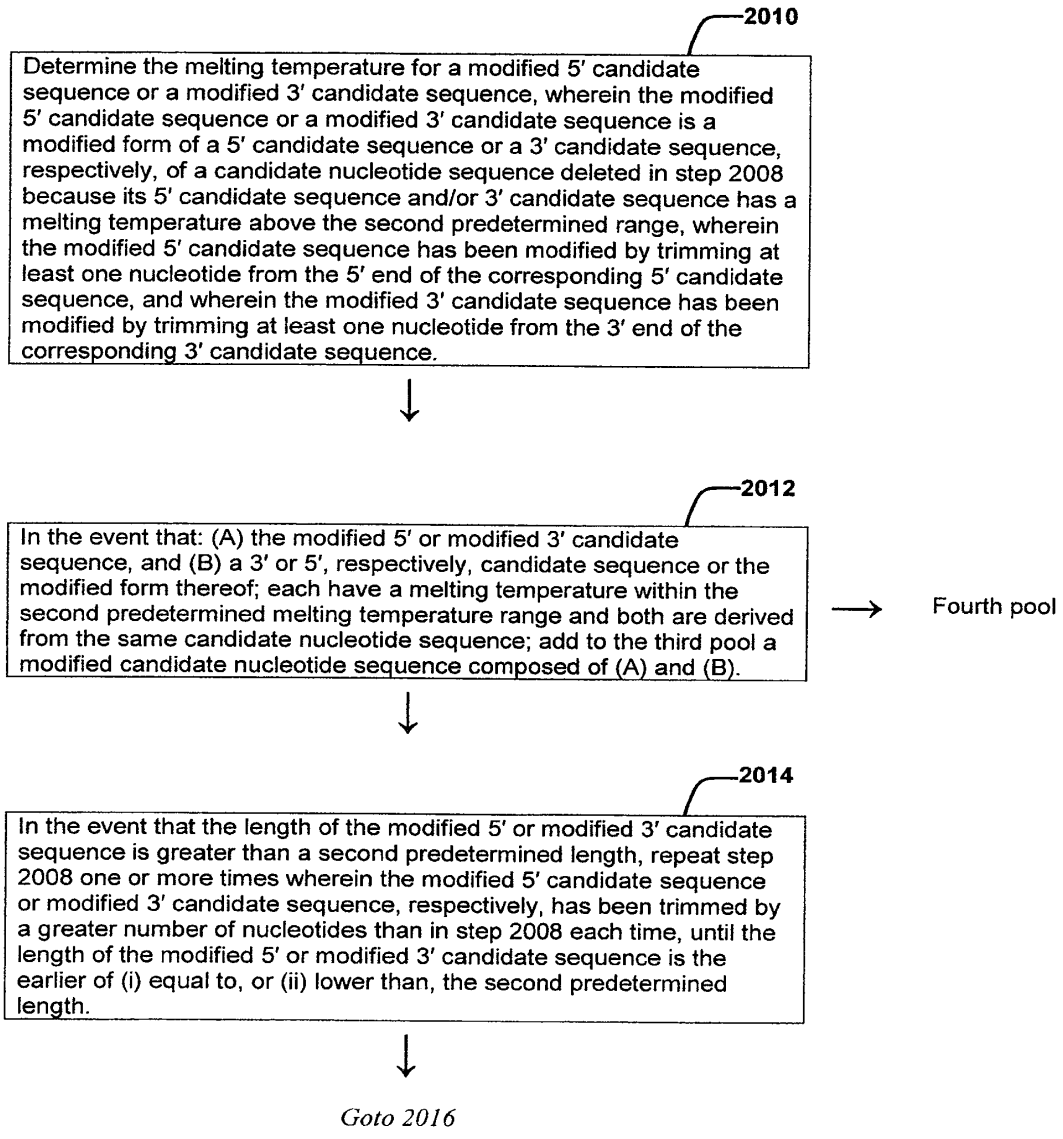

2010
Determine the melting temperature for a modified 5′ candidate sequence or a modified 3′ candidate sequence, wherein the modified 5′ candidate sequence or a modified 3′ candidate sequence is a modified form of a 5′ candidate sequence or a 3′ candidate sequence, respectively, of a candidate nucleotide sequence deleted in step 2008 because its 5′ candidate sequence and/or 3′ candidate sequence has a melting temperature above the second predetermined range, wherein the modified 5′ candidate sequence has been modified by trimming at least one nucleotide from the 5′ end of the corresponding 5′ candidate sequence, and wherein the modified 3′ candidate sequence has been modified by trimming at least one nucleotide from the 3′ end of the corresponding 3′ candidate sequence.
2012
In the event that: (A) the modified 5′ or modified 3′ candidate sequence, and (B) a 3′ or 5′, respectively, candidate sequence or the modified form thereof; each have a melting temperature within the second predetermined melting temperature range and both are derived from the same candidate nucleotide sequence; add to the third pool a modified candidate nucleotide sequence composed of (A) and (B).
Fourth pool
2014
In the event that the length of the modified 5′ or modified 3′ candidate sequence is greater than a second predetermined length, repeat step 2008 one or more times wherein the modified 5′ candidate sequence or modified 3′ candidate sequence, respectively, has been trimmed by a greater number of nucleotides than in step 2008 each time, until the length of the modified 5′ or modified 3′ candidate sequence is the earlier of (i) equal to, or (ii) lower than, the second predetermined length.
Goto 2016

FIGURE 20B

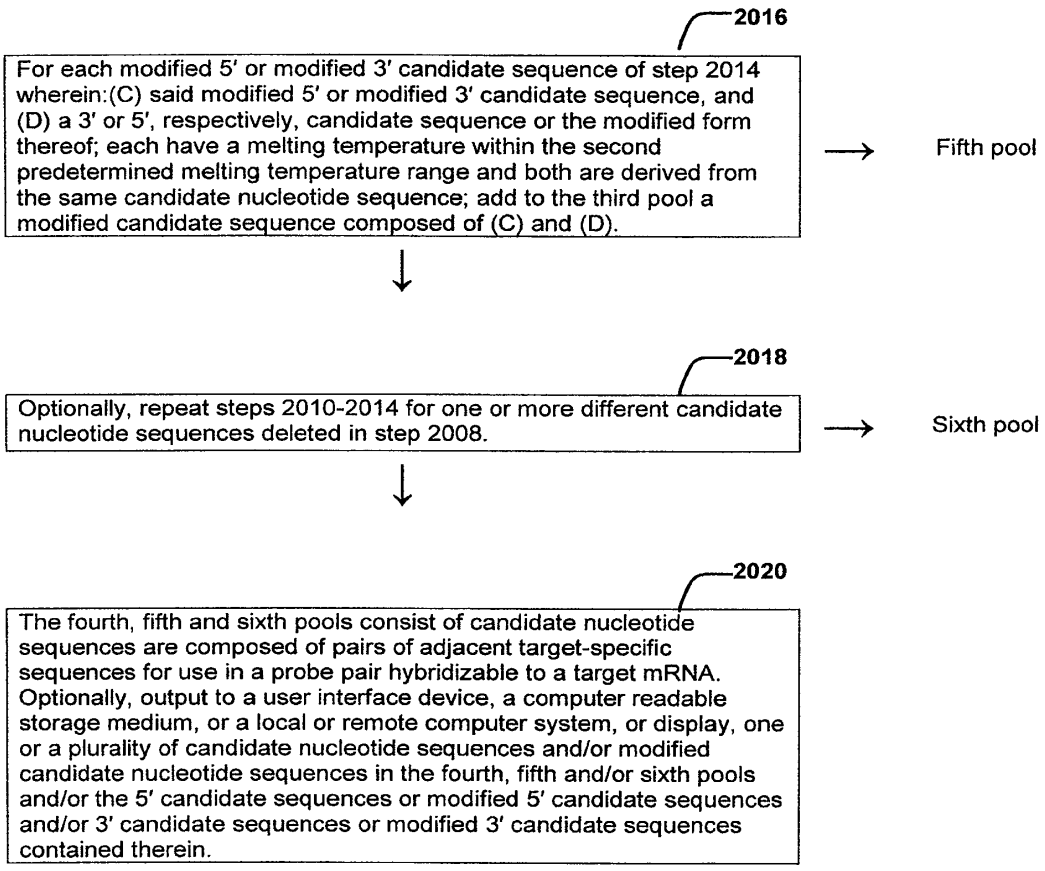

2016
For each modified 5' or modified 3' candidate sequence of step 2014 wherein:(C) said modified 5' or modified 3' candidate sequence, and (D) a 3' or 5', respectively, candidate sequence or the modified form thereof; each have a melting temperature within the second predetermined melting temperature range and both are derived from the same candidate nucleotide sequence; add to the third pool a modified candidate sequence composed of (C) and (D).
Fifth pool
2018
Optionally, repeat steps 2010-2014 for one or more different candidate nucleotide sequences deleted in step 2008.
Sixth pool
2020
The fourth, fifth and sixth pools consist of candidate nucleotide sequences are composed of pairs of adjacent target-specific sequences for use in a probe pair hybridizable to a target mRNA. Optionally, output to a user interface device, a computer readable storage medium, or a local or remote computer system, or display, one or a plurality of candidate nucleotide sequences and/or modified candidate nucleotide sequences in the fourth, fifth and/or sixth pools and/or the 5' candidate sequences or modified 5' candidate sequences and/or 3' candidate sequences or modified 3' candidate sequences contained therein.

FIGURE 20C

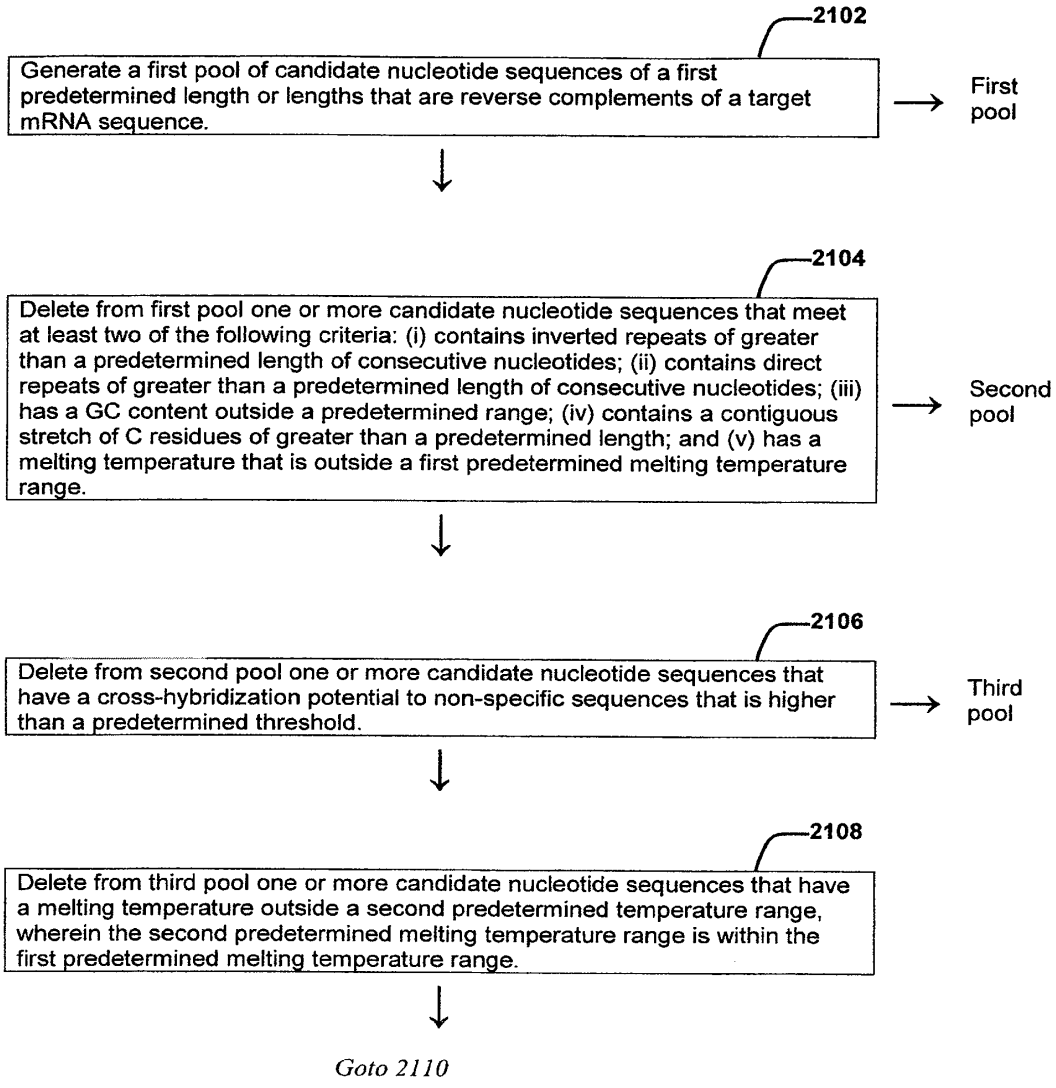

2102
Generate a first pool of candidate nucleotide sequences of a first predetermined length or lengths that are reverse complements of a target mRNA sequence.
First pool
2104
Delete from first pool one or more candidate nucleotide sequences that meet at least two of the following criteria: (i) contains inverted repeats of greater than a predetermined length of consecutive nucleotides; (ii) contains direct repeats of greater than a predetermined length of consecutive nucleotides; (iii) has a GC content outside a predetermined range; (iv) contains a contiguous stretch of C residues of greater than a predetermined length; and (v) has a melting temperature that is outside a first predetermined melting temperature range.
Second pool
2106
Delete from second pool one or more candidate nucleotide sequences that have a cross-hybridization potential to non-specific sequences that is higher than a predetermined threshold.
Third pool
2108
Delete from third pool one or more candidate nucleotide sequences that have a melting temperature outside a second predetermined temperature range, wherein the second predetermined melting temperature range is within the first predetermined melting temperature range.
Goto 2110

FIGURE 21A

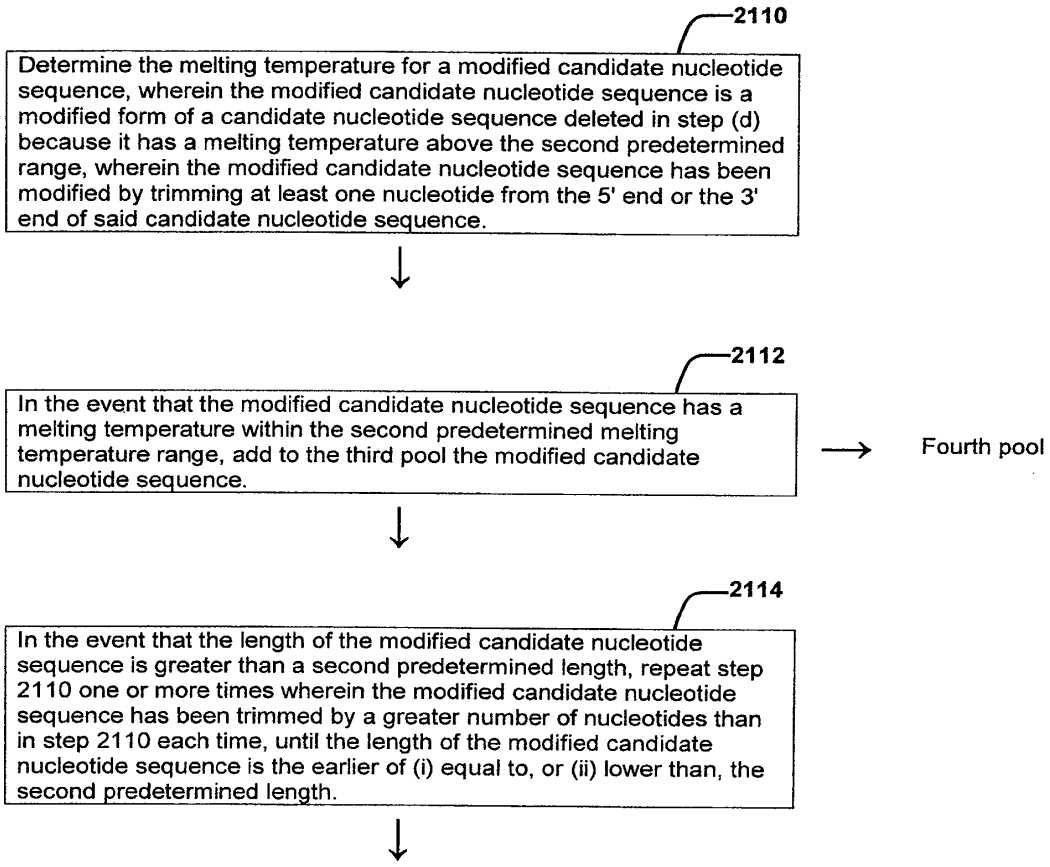

2110
Determine the melting temperature for a modified candidate nucleotide sequence, wherein the modified candidate nucleotide sequence is a modified form of a candidate nucleotide sequence deleted in step (d) because it has a melting temperature above the second predetermined range, wherein the modified candidate nucleotide sequence has been modified by trimming at least one nucleotide from the 5' end or the 3' end of said candidate nucleotide sequence.
2112
In the event that the modified candidate nucleotide sequence has a melting temperature within the second predetermined melting temperature range, add to the third pool the modified candidate nucleotide sequence.
Fourth pool
2114
In the event that the length of the modified candidate nucleotide sequence is greater than a second predetermined length, repeat step 2110 one or more times wherein the modified candidate nucleotide sequence has been trimmed by a greater number of nucleotides than in step 2110 each time, until the length of the modified candidate nucleotide sequence is the earlier of (i) equal to, or (ii) lower than, the second predetermined length.
Goto 2116

FIGURE 21B

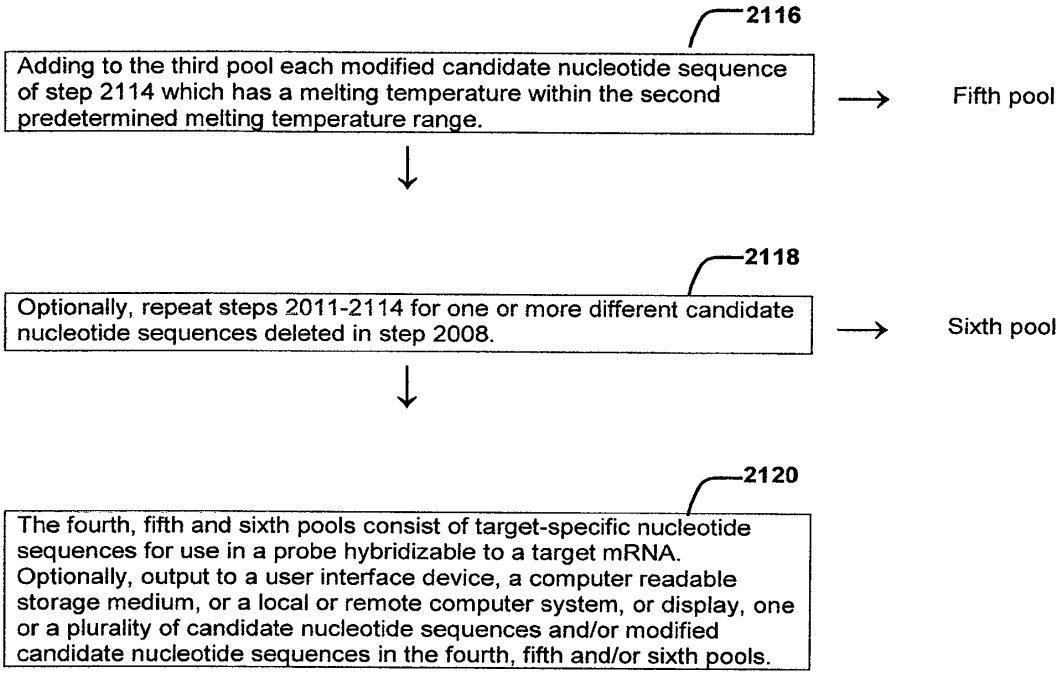

2116
Adding to the third pool each modified candidate nucleotide sequence of step 2114 which has a melting temperature within the second predetermined melting temperature range.
Fifth pool
2118
Optionally, repeat steps 2011-2114 for one or more different candidate nucleotide sequences deleted in step 2008.
Sixth pool
2120
The fourth, fifth and sixth pools consist of target-specific nucleotide sequences for use in a probe hybridizable to a target mRNA. Optionally, output to a user interface device, a computer readable storage medium, or a local or remote computer system, or display, one or a plurality of candidate nucleotide sequences and/or modified candidate nucleotide sequences in the fourth, fifth and/or sixth pools.

FIGURE 21C

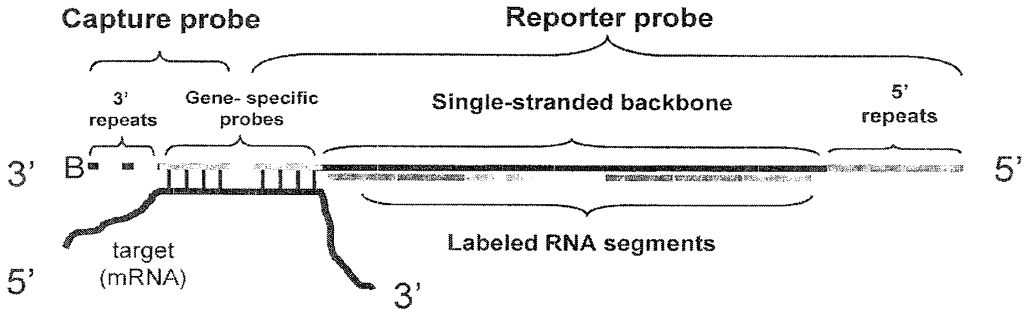


FIGURE 22A

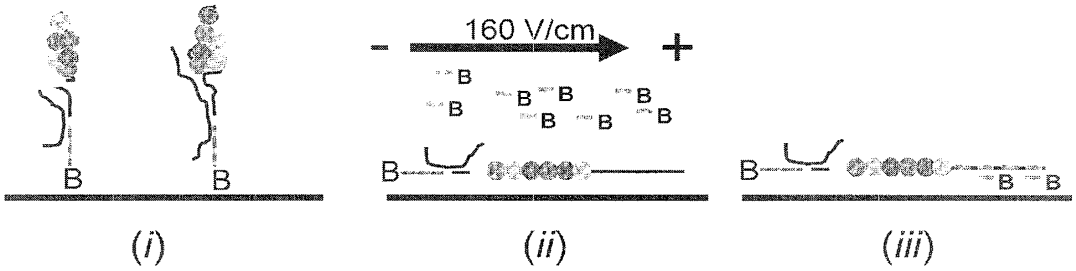


FIGURE 22B

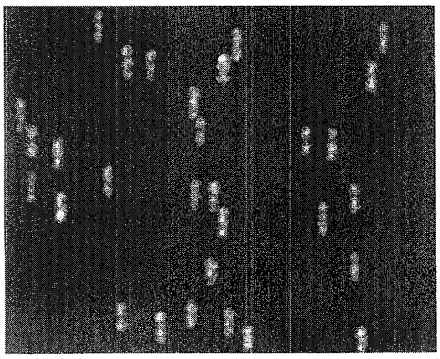


FIGURE 22C

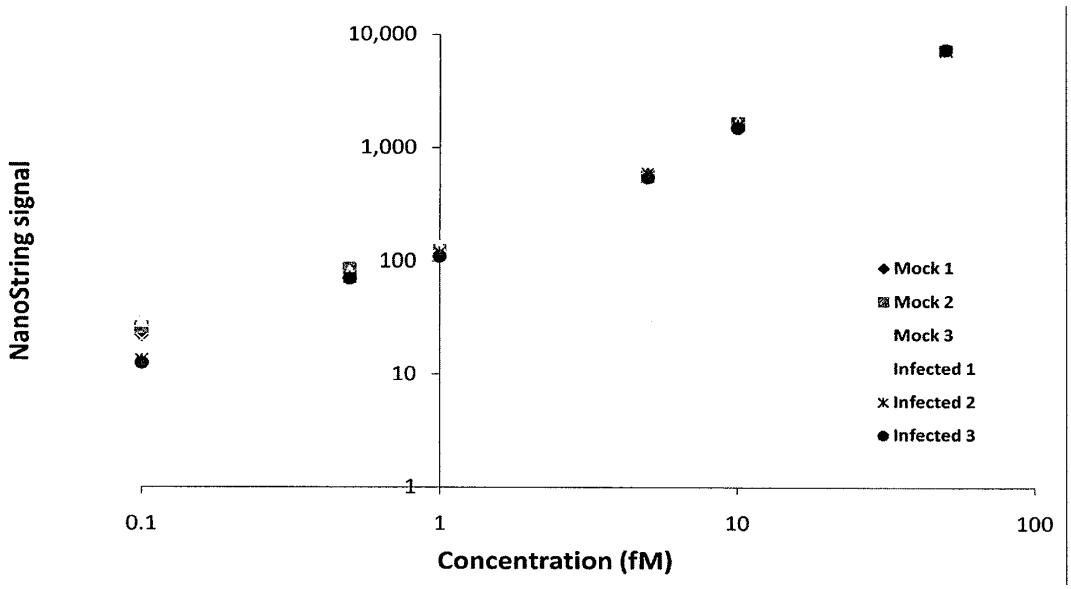


FIGURE 23A

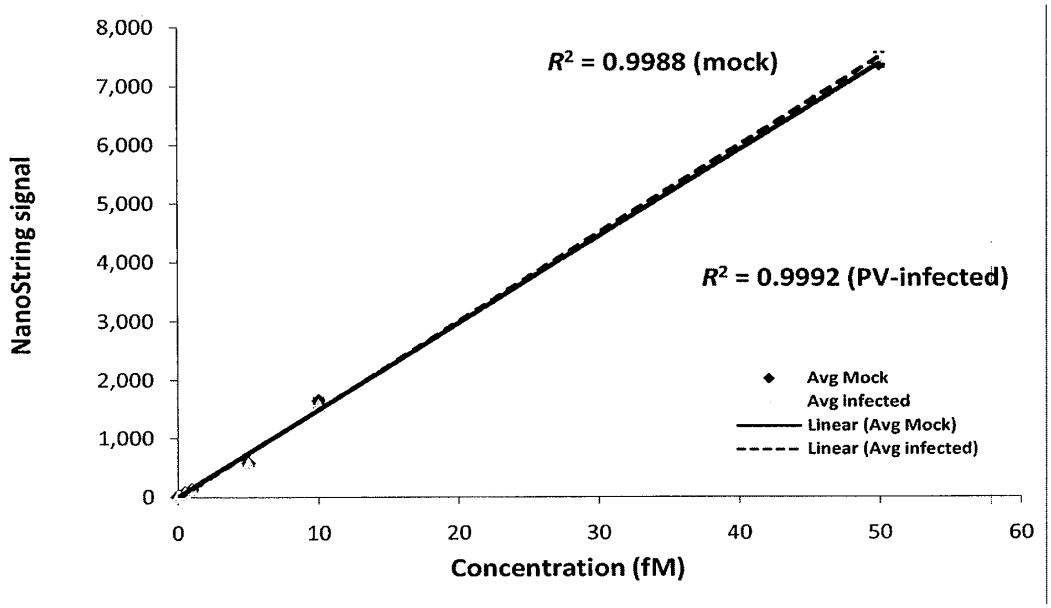


FIGURE 23B

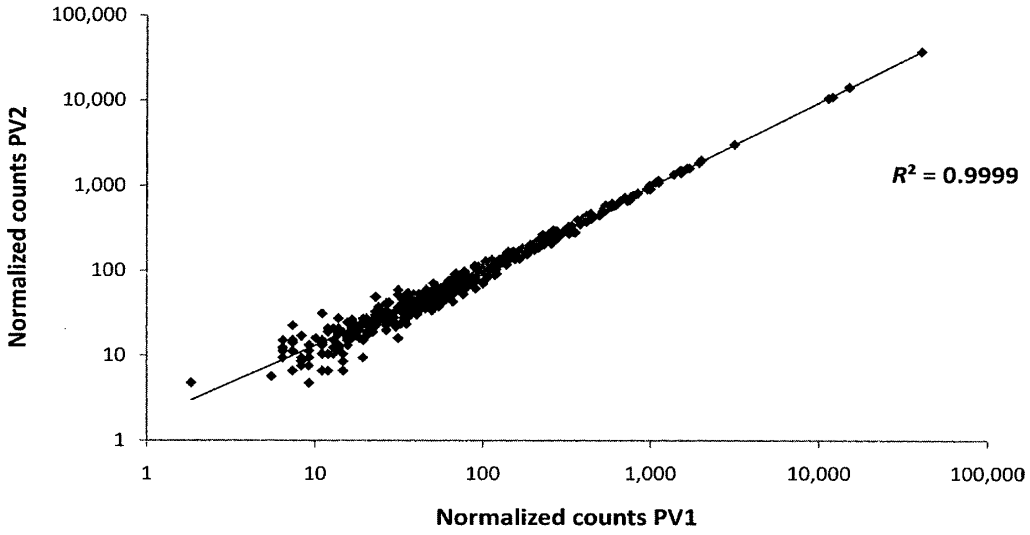


FIGURE 24A

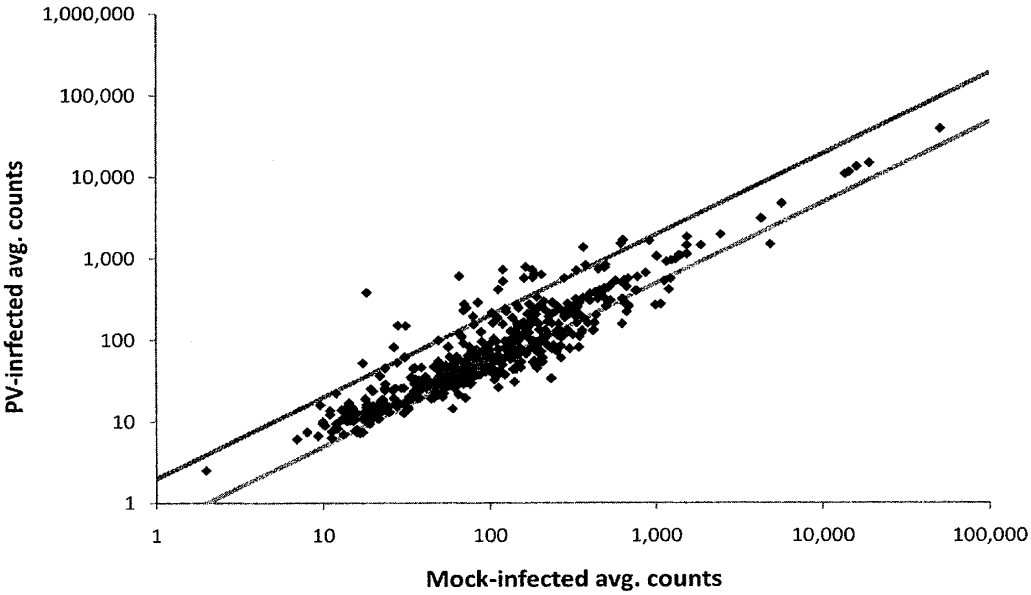


FIGURE 24B

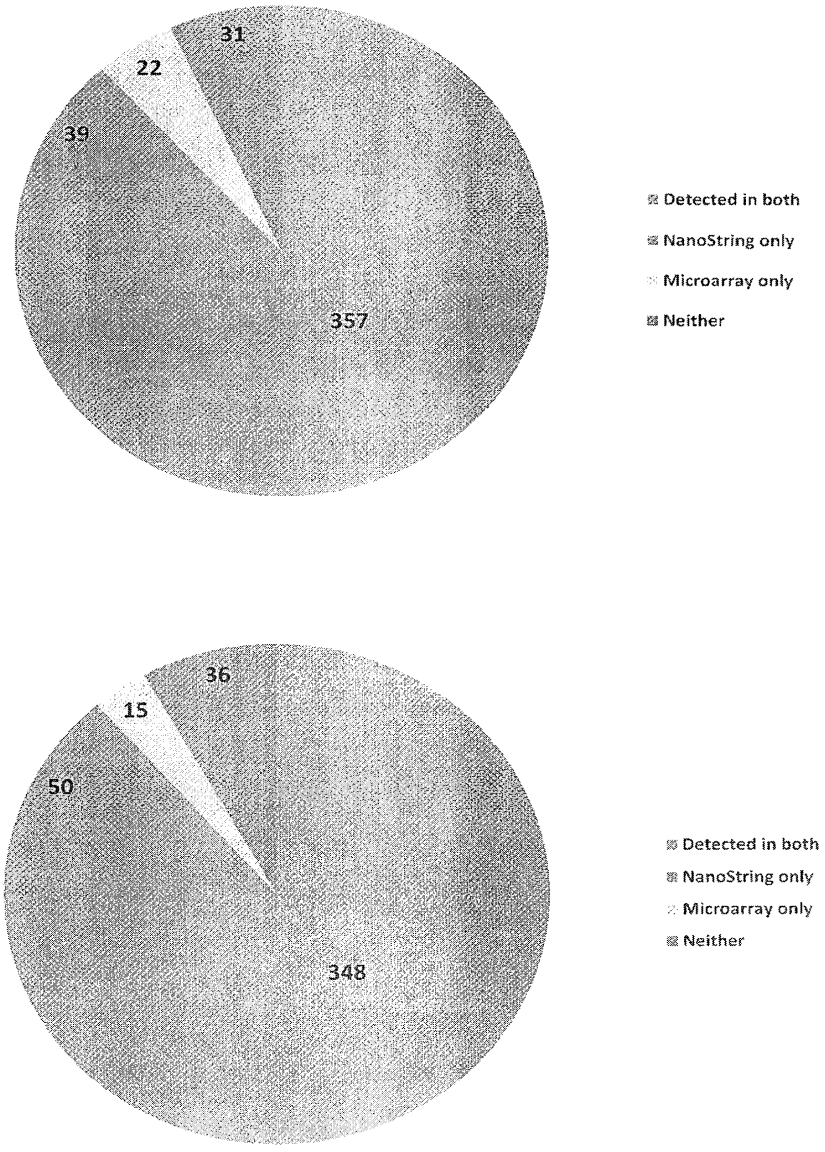

31
22
39
357
Detected in both
NanoString only
Microarray only
Neither
36
15
50
348
Detected in both
NanoString only
Microarray only
Neither

FIGURE 25

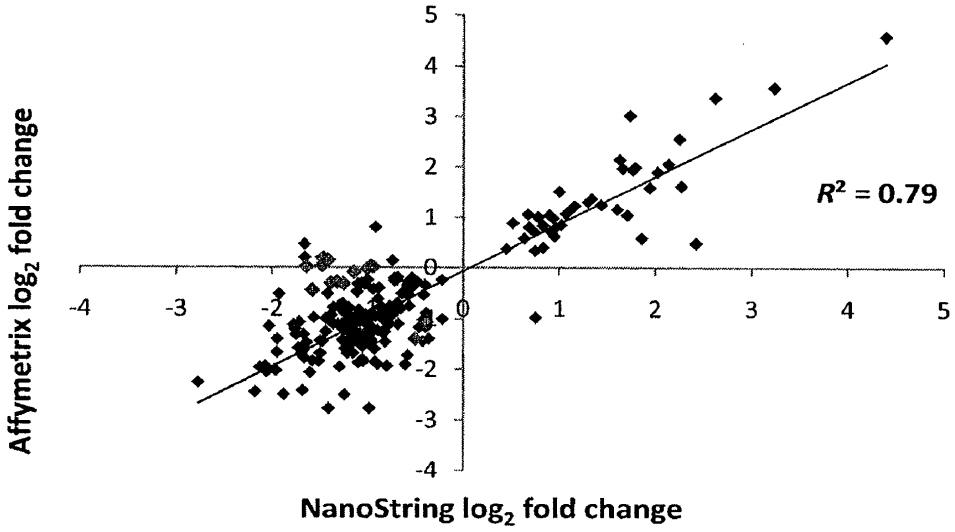


FIGURE 26A

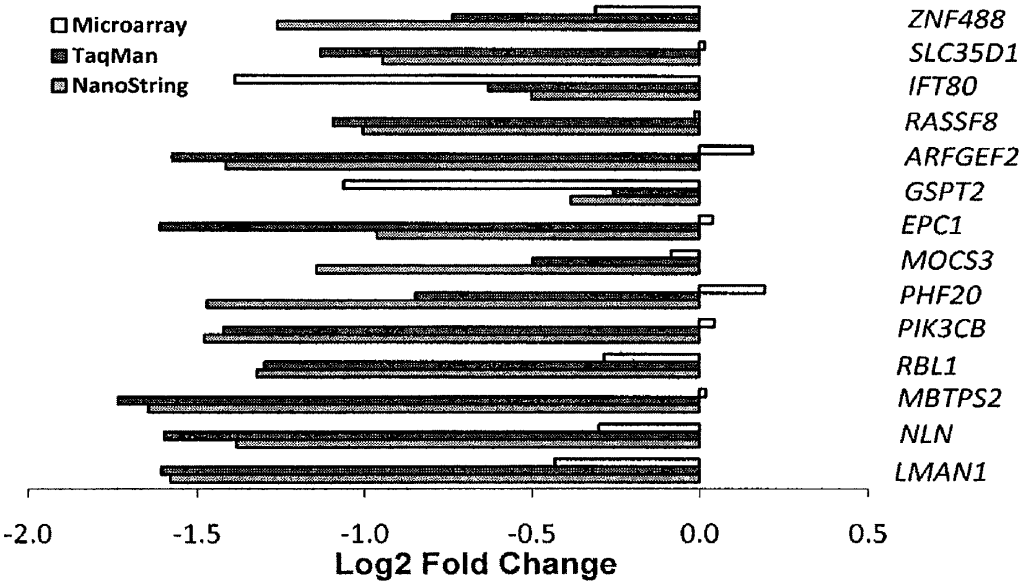


FIGURE 26B

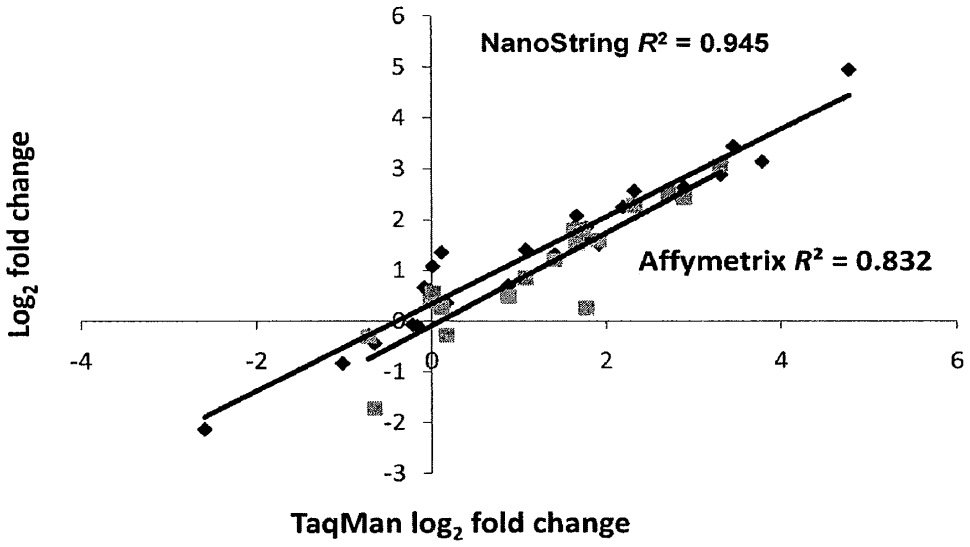
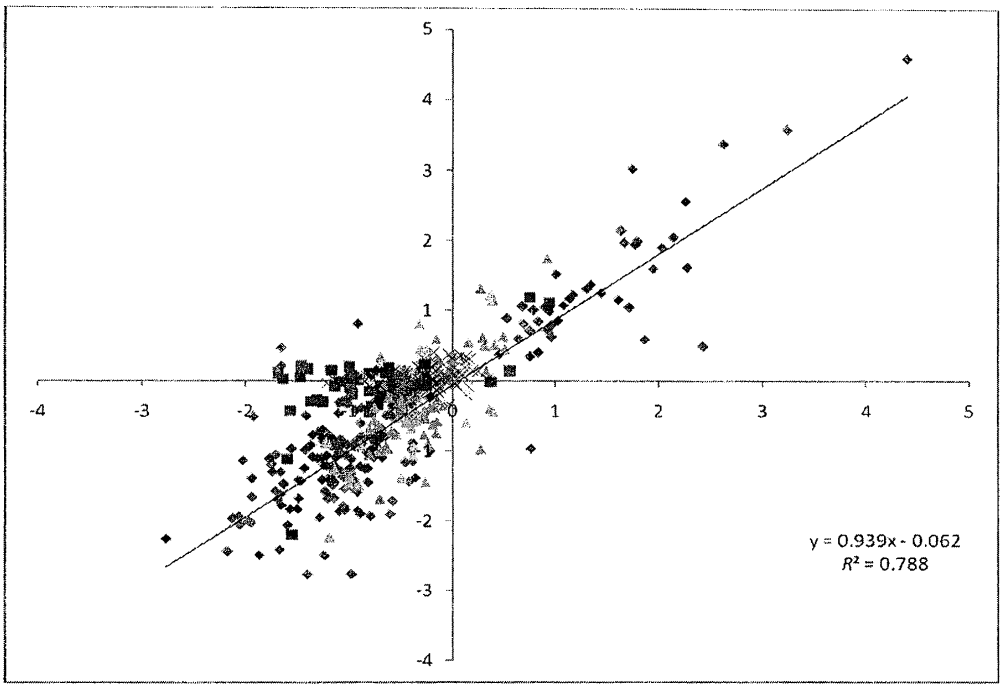

NanoString R^2 = 0.945
Affymetrix R^2 = 0.832
Log_2 fold change
TaqMan log_2 fold change
6
5
4
3
2
1
0
-1
-2
-3
-4
-2
0
2
4
6

FIGURE 26C



Supplementary Figure 1: ***Comparison of fold change results for all 509 genes***

Scatter plot of log2 fold change for 317 genes that were measured by both NanoString and Affymetrix platforms. Genes are color-coded based on the significance of their fold change values (P≤0.05) in either both platforms (♦), NanoString platform only (■), Affymetrix platform only (▲), or neither platform (X). The *R*2 value shown represents the correlation of fold changes of genes that were found to be significant in both NanoString and microarray platforms.

FIGURE 27

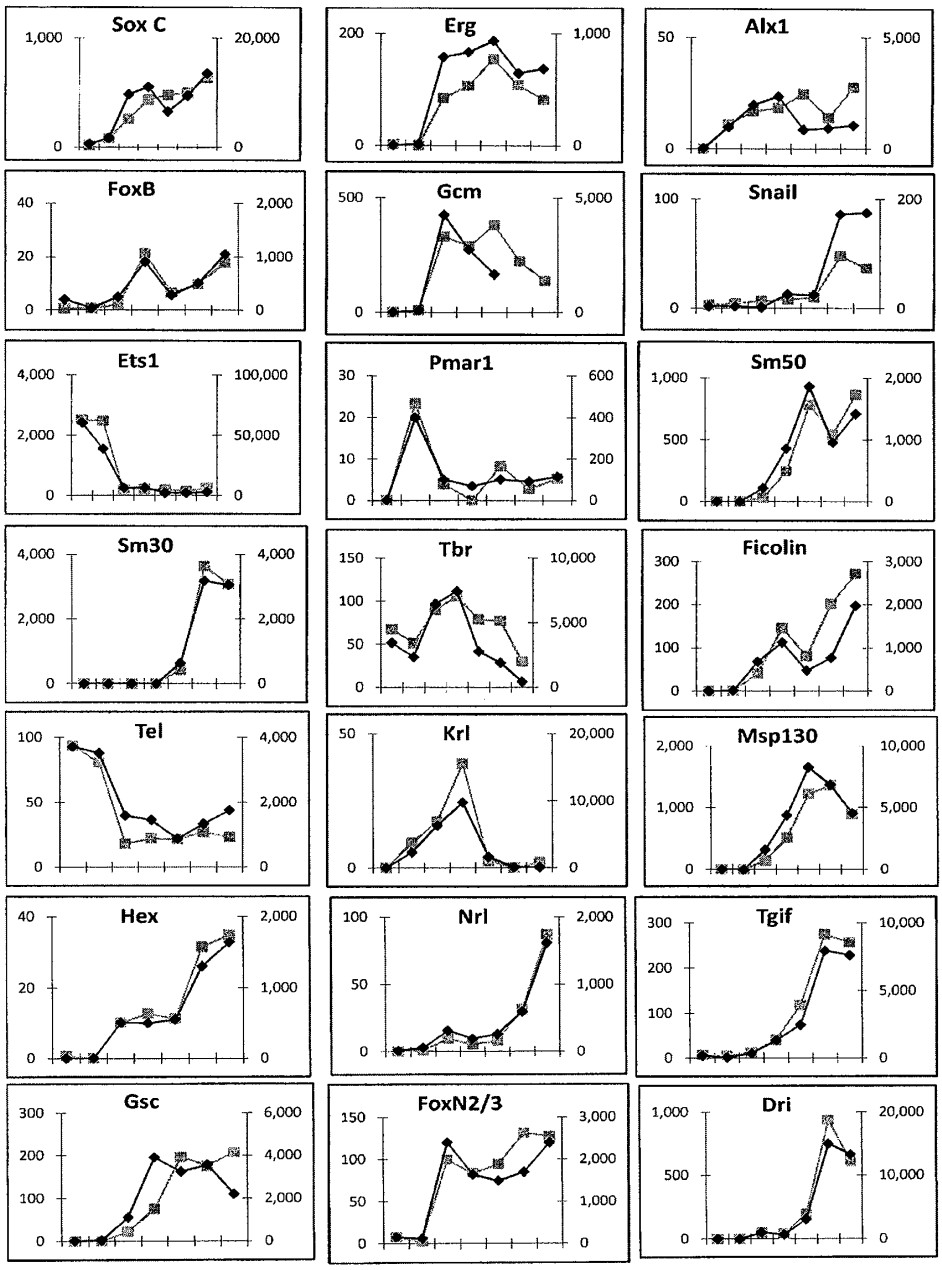

Sox C
1,000
0
20,000
0
Erg
200
0
1,000
0
Alx1
50
0
5,000
0
FoxB
40
20
0
2,000
1,000
0
Gcm
500
0
5,000
0
Snail
100
0
200
0
Ets1
4,000
2,000
0
100,000
50,000
0
Pmar1
30
20
10
0
600
400
200
0
Sm50
1,000
500
0
2,000
1,000
0
Sm30
4,000
2,000
0
4,000
2,000
0
Tbr
150
100
50
0
10,000
5,000
0
Ficolin
300
200
100
0
3,000
2,000
1,000
0
Tel
100
50
0
4,000
2,000
0
Krl
50
0
20,000
10,000
0
Msp130
2,000
1,000
0
10,000
5,000
0
Hex
40
20
0
2,000
1,000
0
Nrl
100
50
0
2,000
1,000
0
Tgif
300
200
100
0
10,000
5,000
0
Gsc
300
200
100
0
6,000
4,000
2,000
0
FoxN2/3
150
100
50
0
3,000
2,000
1,000
0
Dri
1,000
500
0
20,000
10,000
0

FIGURE 28

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 October 2008 (16.10.2008)

PCT

(10) International Publication Number
WO 2008/124847 A3

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2008/059959

(22) International Filing Date: 10 April 2008 (10.04.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/922,817 10 April 2007 (10.04.2007) US
61/029,220 15 February 2008 (15.02.2008) US

(71) Applicant *(for all designated States except US)*: **NANOSTRING TECHNOLOGIES, INC.** [US/US]; 201 Elliott Avenue West, Suite 300, Seattle, Washington 98119 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **GEISS, Gary, K.** [US/US]; 201 Elliott Avenue West, Suite 300, Seattle, Washington 98119 (US). **DAHL, Timothy** [US/US]; 201 Elliott Avenue, Suite 300, Seattle, Washington 98119 (US). **DAVIDSON, Eric, H.** [US/US]; 201 Elliott Avenue, Suite 300, Seattle, Washington 98119 (US).

(74) Agents: **LAHERTY, Carol, D.** et al.; Seed Intellectual Property Law Group PLLC, Suite 5400, 701 Fifth Avenue, Seattle, Washington 98104-7064 (US).

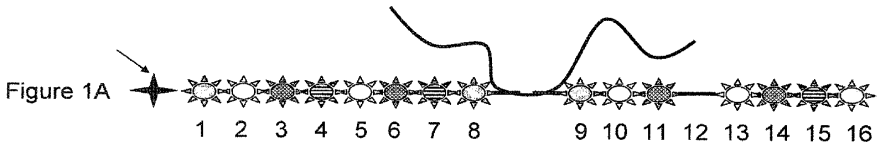
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, 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, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, 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:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
19 February 2009

(54) Title: METHODS AND COMPUTER SYSTEMS FOR IDENTIFYING TARGET-SPECIFIC SEQUENCES FOR USE IN NANOREPORTERS



(57) Abstract: The present invention relates to compositions and methods for detection and quantification of individual target molecules in biomolecular samples. In particular, the invention relates to coded, labeled probes that are capable of binding to and identifying target molecules based on the probes' label codes. Methods, computers, and computer program products for identifying target-specific sequences for inclusion in the probes are also provided, as are methods of making and using such probes. The probes can be used in diagnostic, prognostic, quality control and screening applications.

WO 2008/124847 A3

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1104

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2008/059959

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68

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)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, BIOSIS, WPI Data, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LI F ET AL: "SELECTION OF OPTIMAL DNA OLIGOS FOR GENE EXPRESSION ARRAYS" BIOINFORMATICS, OXFORD UNIVERSITY PRESS, SURREY, GB, vol. 17, no. 11, 1 January 2001 (2001-01-01), pages 1067-1076, XP001062521 ISSN: 1367-4803	1-6, 12-55
Y	the whole document	7-11
Y	US 2003/232382 A1 (BRENNAN THOMAS M [US] ET AL) 18 December 2003 (2003-12-18) paragraphs [0250] - [0258]	7-11
A	WO 02/29379 A (CELADON LAB INC [US]; PETERSON RAYMOND J [US]) 11 April 2002 (2002-04-11) the whole document	1

-/--

[X] Further documents are listed in the continuation of Box C.

[X] 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 document 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 November 2008	11/12/2008

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Luzzatto, Enrico

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2008/059959

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	HE ZHILI ET AL: "Empirical establishment of oligonucleotide probe design criteria" APPLIED AND ENVIRONMENTAL MICROBIOLOGY, AMERICAN SOCIETY FOR MICROBIOLOGY, US, vol. 71, no. 7, 1 July 2005 (2005-07-01), pages 3753-3760, XP002401320 ISSN: 0099-2240 discussion; table 1	1
A	WO 2004/099443 A (FEBIT AG [DE]; DAHMS MICHAEL [DE]; SCHLAUERSBACH ANDREA [DE]; BAUM MIC) 18 November 2004 (2004-11-18) claims	1
A	LOCKHART D J ET AL: "EXPRESSION MONITORING BY HYBRIDIZATION TO HIGH-DENSITY OLIGONUCLEOTIDE ARRAYS" BIO/TECHNOLOGY, NATURE PUBLISHING CO. NEW YORK, US, vol. 14, no. 13, 1 December 1996 (1996-12-01), pages 1675-1680, XP002022521 ISSN: 0733-222X the whole document	1,7,13, 50
P,X	GEISS GARY K ET AL: "Direct multiplexed measurement of gene expression with color-coded probe pairs" NATURE BIOTECHNOLOGY, vol. 26, no. 3, March 2008 (2008-03), pages 317-325, XP002505107 ISSN: 1087-0156 the whole document	1-55

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2008/059959

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2003232382 A1	18-12-2003	NONE	
WO 0229379 A	11-04-2002	AU 1141602 A	15-04-2002
WO 2004099443 A	18-11-2004	EP 1620823 A2	01-02-2006
		US 2006241870 A1	26-10-2006

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1107

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 February 2010 (18.02.2010)

PCT

(10) International Publication Number
WO 2010/019826 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2009/053790

(22) International Filing Date:
13 August 2009 (13.08.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/088,988 14 August 2008 (14.08.2008) US

(71) Applicant *(for all designated States except US)*: **NANOSTRING TECHNOLOGIES, INC** [US/US]; 530 Fairview Ave N, Ste 2000, Seattle, WA 98109 (US).

(71) Applicant *(for US only)*: **DAHL, Craig, E.** [US/US]; Legal Representative Of, Dahl, Tim (deceased), 2094 Jordan Avenue, Juneau, AR 99801 (US).

(72) Inventor: **DAHL, Tim** *(deceased)*.

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **WEBSTER, Philippa, J.** [US/US]; 4317 Desmore Ave. North, Seattle, WA 98103 (US).

(74) Agents: **BORCHARDT, Paul, E.** et al.; Wilson Sonsini Goodrich & Rosati, 650 Page Mill Road, Palo Alto, CA 94304-1050 (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, 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, 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, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, 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, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: STABLE NANOREPORTERS

(57) Abstract: The present invention relates to compositions and methods for detection and quantification of individual target molecules in biomolecular samples. In particular, the invention relates to improved, stable nanoreporter probes that are capable of binding to and identifying target molecules based on the probes' uniquely detectable signal. Methods for identifying target-specific sequences for inclusion in the probes are also provided, as are methods of making and using such probes. Polynucleotide sequences of certain nanoreporter components are also provided. The probes can be used in diagnostic, prognostic, quality control and screening applications.

WO 2010/019826 A1

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1108

STABLE NANOREPORTERS

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 61/088,988, filed August 14, 2008, which application is incorporated herein by reference in its entirety.

BACKGROUND OF THE INVENTION

[0002] Although all cells in the human body contain the same genetic material, the same genes are not active in all of those cells. Alterations in gene expression patterns can have profound effects on biological functions. These variations in gene expression can be at the core of altered physiologic and pathologic processes. Therefore, identifying and quantifying the expression of genes in cells can aid the discovery of new therapeutic and diagnostic targets.

[0003] To date there are several techniques available that allow the detection of the expression level of multiple genes in a complex sample at one time. Most of these technologies employ, DNA microarrays, devices that consist of thousands of immobilized DNA sequences present on a miniaturized surface that have made this process more efficient. Unfortunately, despite the miniaturization of array formats, this method still requires significant amounts of the biological sample. However, in several cases, such as biopsies of diseased tissues or samples of a discrete cell type, the biological sample is in limited supply. In addition, the kinetics of hybridization on the surface of a microarray is less efficient than hybridization in small amounts of aqueous solution. Moreover, while methods exist to estimate the amount of nucleic acid present in a sample based on microarray hybridization result, microarray technology thus far does not allow for detection of target molecules on an individual level, nor are there microarray-based methods for directly quantifying the amount of target molecule in a given sample.

[0004] Thus, there exists a need for accurate and sensitive detection, identification and quantification of target molecules in complex mixtures.

SUMMARY OF THE INVENTION

[0005] This invention relates generally to the field of detection, identification, and quantification of target molecules in a sample. The present invention relates in part to improved, stable nanoreporters, and populations of these nanoreporters, based in part on certain design features in the polynucleotide sequences of the nanoreporter backbones and complementary polynucleotide sequences attached thereto.

[0006] In some embodiments, the invention provides a population of uniquely labeled nanoreporter probes, where each nanoreporter probe comprises: i) an unique target-specific region; and ii) a region comprising an unique, designed nanoreporter wherein the nanoreporter comprises a single-stranded

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1109

nucleic acid backbone, the backbone comprising a plurality of label attachment regions covalently attached together in an unique linear combination, where the label attachment regions are selected from a population of designed polynucleotide sequences, where each polynucleotide sequence is hybridized to a complementary polynucleotide sequence having attached thereto one or more detectable molecules; where each complementary polynucleotide sequence has designated a specific detectable molecule; and where each nanoreporter has a detectable signal that distinguishes it from other nanoreporters in said population. In some embodiments, each label attachment region of each backbone is different from the other label attachment regions in that same backbone. In some embodiments, each of the nanoreporter probes further comprises a constant region, wherein the constant region comprises a plurality of repeat nucleotide sequences. The population of nanoreporter probes can comprise two or more nanoreporter probes.

[0007] In some embodiments, the invention provides methods for determining the presence of at least one target molecule in a sample, comprising: (1) forming at least one molecular complex comprising: (a) at least one target molecule, and (b) at least one probe comprising an unique target-specific region and a region comprising an unique, designed nanoreporter, wherein said nanoreporter comprises a single-stranded nucleic acid backbone, the backbone comprising a plurality of label attachment regions covalently attached together in an unique linear combination, wherein each label attachment region is hybridized to a complementary polynucleotide sequence having attached thereto one or more detectable molecules, wherein each complementary polynucleotide sequence has designated a specific detectable molecule; and individually counting the presence of one or more molecular complex or at least part of the at least one molecular complex to determine the presence of said at least one target molecule in the sample. In some embodiments, the percentage of valid molecular counts of the molecular complex is higher that about 12.5%. In some embodiments, the numbers of counts is at least two fold higher than the counts obtain when using a nanoreporter probe comprising M13 DNA. A nanoreporter comprising M13 DNA can comprises a single–stranded backbone comprising a plurality of M13 DNA regions covalently attached together wherein each region is hybridized to a complementary polynucleotide sequence having attached thereto one or more detectable molecules.

[0008] In some embodiments of the methods and composition of the invention, the numbers of counts above background of each of the molecular complex after normalization of the sample is at least two fold higher when compared to a nanoreporter probe comprising M13 DNA. In some embodiments, the Tm of the complementary polynucleotide sequences when hybridized to its label attachment regions is about 80°C or higher. In some embodiments, the melting temperature (Tm) of the complementary polynucleotide sequences when hybridized to its label attachment regions is higher than the Tm of polynucleotide sequences complementary to M13 DNA when hybridized to the nanoreporter probe comprising M13 DNA.

[0009] In some embodiments, the methods of the invention further comprise determining the presence of a plurality of target molecules by a method comprising forming a plurality molecular

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1110

complexes, each complex comprising (a) at least one target molecule and (b) at least one probe comprising an unique target-specific region and a region comprising an unique, designed nanoreporter, wherein each nanoreporter comprises a single-stranded nucleic acid backbone, the backbone comprising a plurality of label attachment regions covalently attached together in an unique linear combination, wherein each label attachment region is hybridized to a complementary polynucleotide sequence having attached thereto one or more detectable molecules, and wherein each probe comprises a different nanoreporter region. In some embodiments, the presence of at least 5, 10, 20, 30, 50, 100, 200, 300, or 500 different target molecules is determined. In some embodiments, the target molecule is a nucleic acid. In some embodiments, the nucleic acid comprises at least one heritable mutation, at least one somatic mutation, at least one single nucleotide polymorphism (SNP), at least one point mutation, at least one deletion mutation, at least one insertion mutation, at least one chromosomal translocation, or combinations thereof. In some embodiments, the target molecule is a diagnostic indicator.

[0010] In some embodiments, the invention provides, an uniquely labeled nanoreporter probe comprising: i) an unique target-specific region; and ii) a region comprising a plurality of designed label attachment regions covalently attached together in a linear combination, wherein each label attachment region comprises about 800 to 1300 nucleotide bases and has a G/C content of about 50%, wherein each selected label attachment region is different from the other selected label attachment regions, and wherein each label attachment region is hybridized to a complementary polynucleotide sequence having attached thereto one or more detectable molecules, wherein the complementary polynucleotide sequence has a G/C ratio of at least 1/1. In some embodiments, the complementary polynucleotide sequence has a G/C ratio of about 3/2. In some embodiments, the nanoreporter probe further comprises a constant region, where the constant region comprises a plurality of repeat nucleotide sequences.

[0011] In some embodiments of the methods and compositions of the invention, the label attachment region comprises about 800 to 1300 nucleotide bases and has a G/C content of about 50% and the complementary polynucleotide sequence has a G/C ratio of about 1/1. In some embodiments, the complementary polynucleotide sequence has a G/C ratio of about 3/2. Examples of templates that can be used to generate the label attachments regions described herein include, but are not limited to, sequences selected from the group consisting of SEQ ID NOS: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, and 24.

[0012] In some embodiments of the methods and compositions of the invention, the label attachment regions comprise a similar adenine content. In some embodiments, the adenine bases are spaced at least an average of every 8 to 16 nucleotide bases. In some embodiments, the label attachment regions comprise a regularly repeated pattern of adenine bases. In some embodiments, the adenine bases are spaced about every 8 to 16 nucleotide bases. In some embodiments, the label attachment regions comprise a thymine content of about 35-45%.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1111

[0013] In some embodiments of the methods and compositions of the invention, the complementary polynucleotide sequence comprises a RNA polynucleotide sequence. The RNA polynucleotide sequence can comprises at least one amino-allyl modified uracil base. In some embodiments, the detectable molecule in the complementary polynucleotide sequence is attached to the amino-allyl modified uracil base. In some embodiments, the RNA polynucleotide sequence comprises a plurality of amino-allyl modified uracil bases that are spaced at about an average of every 8 to 16 bases in said RNA polynucleotide sequence. In some embodiments, the detectable molecule is attached to each of the allyl modified uracil bases.

[0014] In some embodiments of the compositions and methods of the invention, the detectable molecules are fluorescent dyes.

[0015] In some embodiments, the invention provides methods for preparing at least one uniquely labeled nanoreporter comprising: i) combining a plurality of label attachment region each comprising about 800 to 1300 nucleotide bases and a G/C content of about 50%, wherein each selected label attachment region is different from the other selected label attachment regions, ii) covalently attaching the plurality of label attachment regions to each other in linear combination; and iii) hybridizing a complementary polynucleotide sequence to said label attachment region, wherein said complementary polynucleotide sequence is having attached thereto one or more detectable molecules. In some embodiments, the methods further comprise preparing a labeled nanoreporter probe by attaching the labeled nanoreporter to a target specific region. In some embodiments, the complementary polynucleotide sequence has a G/C ratio of at least 1/1.

[0016] In some embodiments, the invention provides kits for preparing at least one uniquely labeled nanoreporter comprising: a) at least three label attachment regions each comprising about 800 to 1300 nucleotide bases, a G/C content of about 50%; and b) at least three complementary polynucleotide sequences having attached thereto a detectable molecule, wherein the complementary polynucleotide sequences has a G/C ratio of at least 1/1. In some embodiments, the kits further comprise at least three target specific probes.

[0017] In some embodiments, the invention provides kits comprising a population of nanoreporters as described herein and instructions for its use.

INCORPORATION BY REFERENCE

[0018] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1112

reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0020] Figure 1 is a schematic illustration of a dual nanoreporter with a 6-position nanoreporter code, using a capture probe and a single 6-position nanoreporter component. The arrow is illustrative of an affinity tag, which is optionally included and which can be used to purify the nanoreporter or immobilize the nanoreporter (or nanoreporter-target molecule complex) for the purpose of imaging.

[0021] Figure 2A-2E: Figure 2A is a schematic illustration of the experiment shown in Figures 2B and 2C. In this case, the diamond represents biotin that was used to attach the complex by one end to the surface prior to stretching. Figures 2B and 2C show images from experiments in which S2-A capture probe, S2-B labeled nanoreporter and S2 target DNA (Figure 2B) or S2 target RNA (Figure 2C) were hybridized. Figure 2E shows a close-up of nanoreporter complexes from Figure 2B, each containing S2-A capture probe, S2-B labeled nanoreporter and S2 target DNA. Figure 2D shows an image of a negative control experiment, in which S2-A capture probe, S2-B labeled nanoreporter and no S2 target RNA were hybridized.

[0022] Figure 3A-3B: Shows a labeled (nucleic acid-based) nanoreporter with an affinity tag, A1, at one end. In Figure 3, the labeled nanoreporter is immobilized through the binding of A1 to an immobilized affinity partner. The other end of the nanoreporter is in solution (Figure 3A), but can be immobilized by hybridization to a complementary oligonucleotide which contains another affinity tag (A2) used to immobilize the nanoreporter (Figure 3B). A1 and A2 can be the same, for example biotin, for immobilization on an avidin- or streptavidin-coated surface. Upon immobilization of A1, the nanoreporter can be stretched, or "elongated" as depicted in Figure 3, for example by electrostretching, for separation of the label attachment regions in a manner that permits detection of the nanoreporter code. Optionally, while the nanoreporter is in an elongated state, A2 is introduced and binds the end of the nanoreporter that is complementary to A2 down to the surface.

[0023] Figure 4A-4C. Figure 4A illustrates immobilization of one terminus of a DNA molecule in a microfluidic device; Figure 4B illustrates extension of the DNA in an electric field; and Figure 4C illustrates selective immobilization of a second terminus of the extended DNA molecule by affinity tags introduced into the device following elongation.

[0024] Figure 5 shows the detection of IL-8 in the same sample using both the DV1 and M13 nanoreporter systems. The data was collected from experiments performed according to Example 4.

[0025] Figure 6 shows a comparison of the DV1 and M13 systems in measuring the expression of 40 genes in two samples. The data was collected from two representative experiments performed according to Example 4.

[0026] Figure 7 shows the detection of GusB expression. In this experiment, M13 and DV1 libraries containing 148 identical probes were used to measure gene expression levels in 26 mouse RNA samples. An average 6-fold increase was seen for the absolute number of counts measured by the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1113

DV1 reporters compared to the M13 reporters. The data was collected from representative experiments performed according to Example 5.

[0027] **Figure 8** shows a graph of "valid reporters," which refers to strings of spots that can be interpreted by the image analysis software as a "countable" reporter. The results are described as a percentage of the total number of binding events in a field-of-view. A significantly higher percentage of DV1 reporters are countable, in this experiment leading to a 3-fold gain in data (from 12.5% to 38%).

[0028] **Figure 9** shows the polynucleotide sequence of a plasmid vector that may be utilized to clone, propagate, and generate the single stranded nanoreporter backbones of the present invention.

[0029] **Figure 10** shows the polynucleotide sequences of two templates that were utilized in the dye optimization experiments to generate single-stranded nanoreporter backbones (*see* Example 6). In these template sequences, the regularly-repeated base is thymine, which upon transcription produces a complementary single-stranded nanoreporter backbone having adenine as the regularly-repeated base. Figure 10A shows the polynucleotide sequence of a template having regularly-repeated base at about every 8 nucleotides. Figure 10B shows the polynucleotide sequence of a template having a regularly-repeated base at about every 10 nucleotides.

[0030] **Figure 11** shows the polynucleotide sequences of two templates that were utilized in the dye optimization experiments to generate single-stranded nanoreporter backbones (*see* Example 6). Figure 11A shows the polynucleotide sequence of a template having regularly-repeated base at about every 12 nucleotides. Figure 11B shows the polynucleotide sequence of a template having a regularly-repeated base at about every 14 nucleotides.

[0031] **Figure 12** shows the polynucleotide sequences of two templates that were utilized in the dye optimization experiments to generate single-stranded nanoreporter backbones (*see* Example 6). Figure 12A shows the polynucleotide sequence of a template having regularly-repeated base at about every 16 nucleotides. Figure 12B shows the polynucleotide sequence of a template having a regularly-repeated base at about every 24 nucleotides.

DETAILED DESCRIPTION OF THE INVENTION

[0032] The present invention provides compositions and methods for detection and quantification of individual target molecules in biomolecular samples. In particular, the invention provides stable nanoreporters that are capable of binding individual target molecules, and provide improved detection of target molecules. Through nanoreporters' label codes, the binding of the nanoreporter probes to target molecules results in the identification of the target molecules. Methods of making and using such nanoreporters are also provided. The nanoreporters can be used in a wide variety of applications such as diagnostic, prognostic, quality control and screening applications.

[0033] Certain aspects of the present invention relate to the selection of a library or population of designed (*e.g.*, synthetic sequences) nanoreporters. More specifically, certain optimized sequence

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1114

characteristics provide improved molecular stability of a nanoreporter, as well as improved detection when the nanoreporter is combined with a target-specific sequence. For instance, the methods and compositions of the invention provide nanoreporters comprising unique synthetic backbones that do not produce a secondary structure and produce consistent brightness.

[0034] In some embodiments, the present invention provides a population of designed (e.g. synthetic) nanoreporter wherein said nanoreporter comprises a plurality of different detectable molecules, and wherein the plurality of different detectable molecules in each nanoreporter has a detectable signal that distinguishes it from other nanoreporters in said population. Thus, in some embodiments the present invention provides a population of designed nanoreporters with improved molecular stability, wherein each nanoreporter in the population has a detectable signal that distinguishes it from other nanoreporters in said population.

[0035] In some embodiments, the invention provides nanoreporters comprising unique designed backbones (e.g. synthetic) hybridized to an unique complementary polynucleotide sequence having attached thereto a detectable label. Each nanoreporter will generate an unique signal that will not change during the course of an detection assay. That is the unique and distinguishable signal or signal code associated with the nanoreporter will remain the same during the course of an detection assay. In some embodiments, the complementary polynucleotide sequence having attached thereto the detectable label can be an in-vitro transcribed, dye-coupled RNA segments having consistent brightness. Consistent brightness as described herein refers to the strength, size and/or intensity of the signal produced by the dye-coupled RNA segments. That is, in some embodiments, the strength, size and/or intensity of the signal of the dye will be similar among dye-coupled RNA segments that have the same dye attached thereto. For example, dye-coupled RNA segments that have a color green dye attached thereto will have similar a signal or a signal with same intensity and/or brightness. This is useful, among other things, because it allows an image analysis software/algorithm or user to define what a spot of a particular color should look like in term of brightness and size. In addition, several of the dyes might bleed to the channels of the other dyes, if the strength of a signal is consistent, then the bleed through signal will also be consistent, thus, allowing for the bleed through to be ignored. Furthermore, noises in the images that do not fall within the spot parameters can be disregarded. As a result, for example, one can narrowly define the parameters associated with the nanoreporters' signals allowing the software/algorithm or user to disregard a higher percentage of the noise. These provide for more robust and reliable readings.

[0036] In some embodiments, the nanoreporters comprise a backbone with an arrangement of nucleotide-based label attachment regions, wherein each label attachment region has have a specific sequence designated for a specific label. In this system, the unique sequence of the backbone dictates the color code of the nanoreporter. Each backbone will anneal only to the polynucleotide sequence complementary to its sequence, each of which has a specific, designated label (e.g. color). Thus, each backbone will generate only the designated code, e.g., even if the polynucleotide sequence detach

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1115

from the backbone during the detection process. If the polynucleotide sequence in the nanoreporter detach during synthesis or during a hybridization procedure, they can only be replaced with polynucleotide sequence of the same color, eliminating the potential for shared or swapped labels. The complementary polynucleotide sequence having attached thereto the detectable label can be an in-vitro transcribed, dye-coupled RNA segments. In some embodiments, in constructing a given nanoreporter backbone from a library of individually unique polynucleotide-based label attachment region templates each label attachment region is assigned a detectable label (e.g. a detectable molecule), and each label attachment region within a given backbone is selected to be different from the other label attachment regions in that same backbone.

[0037] In some embodiments, the sequences are designed to have an even distribution of one base, allowing for an even distribution of coupled labels (e.g. fluors) when this base is introduced as a modified nucleotide into an in vitro polymerized RNA or DNA. In some embodiments, the sequences are designed without significant direct or inverted repeats to make the nanoreporters as unstructured and unique as possible. In some embodiments, the sequences are designed without any direct or inverted repeat of 9 nucleotides or greater across a sequence that 1100 base pairs in length. In some embodiments, the sequences are designed without any direct or inverted repeat of 7 nucleotides or greater across any 100 base pair region. In some embodiments, the label attachment regions and/or the complementary polynucleotide sequence comprise a particular G/C content and ratio.

[0038] Examples of polynucleotide templates that may be utilized to generate these designed label attachment regions are set forth in the polynucleotide sequences of SEQ ID NOS:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, and 24.

[0039] Accordingly, certain aspects of the present invention provide a population of unique nanoreporters or nanoreporter probes, each comprised of a unique polynucleotide-based backbone, wherein each nanoreporter in the population is not only distinct from the other nanoreporters in the population but also has improved molecular stability and a detectable signal that distinguishes it from other nanoreporters in said population. In some embodiments each nanoreporter probe comprises a plurality of individual, designed label attachment regions. In some embodiments, each of the label attachment regions is different from the other label attachment regions in that same nanoreporter. Thus, in certain aspects the invention provides a more stable population of unique nanoreporters having improved detection properties. An exemplary nanoreporter having such characteristics is described herein, referred to as the DV1 nanoreporter.

Nanoreporters

[0040] A fully assembled and labeled nanoreporter probe comprises two main portions, a target-specific sequence that is capable of binding to a target molecule, and a labeled nanoreporter which provides a "code" of signals (the "nanoreporter code") associated with the target-specific sequence.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1116

Upon binding of the nanoreporter probe to the target molecule, the nanoreporter code identifies the target molecule to which the nanoreporter is bound.

[0041] Nanoreporters are modular structures. In some embodiments, the nanoreporter comprises a plurality of different detectable molecules. In some embodiments, a labeled nanoreporter, is a molecular entity containing certain basic elements: (i) a plurality of unique label attachment regions attached in a particular, unique linear combination, and (ii) complementary polynucleotide sequences attached to the label attachment regions of the backbone. In some embodiments, the labeled nanoreporter comprises 2, 3, 4, 5, 6, 7, 8, 9, 10 or more unique label attachment regions attached in a particular, unique linear combination, and complementary polynucleotide sequences attached to the label attachment regions of the backbone. In some embodiments, the labeled nanoreporter comprises 3 or more unique label attachment regions attached in a particular, unique linear combination, and complementary polynucleotide sequences attached to the label attachment regions of the backbone. In some embodiments, the labeled nanoreporter comprises 6 or more unique label attachment regions attached in a particular, unique linear combination, and complementary polynucleotide sequences attached to the label attachment regions of the backbone. A nanoreporter probe further comprises a target-specific sequence, also attached to the backbone.

[0042] The term label attachment region includes a region of defined polynucleotide sequence within a given backbone that may serve as an individual attachment point for a detectable molecule. In some embodiments, the label attachment regions comprise designed sequences. Specific examples of defined polynucleotide template sequences that may be utilized to generate single stranded label attachment regions include the template sequences set forth in SEQ ID NOS: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, and 24, in addition to suitable variants thereof (*e.g.*, sequences having 60%, 70%, 80%, 90%, 95%, 98%, or more, including all integers in between, sequence identity). Transcription of the template sequences of SEQ ID NOS:1-24 generates single stranded label attachment regions having a polynucleotide sequence that is complementary to the sequences of SEQ ID NOS:1-24.

[0043] In some embodiments, the label nanoreporter also comprises a backbone containing a constant region. The term constant region includes tandemly-repeated sequences of about 10 to about 25 nucleotides that are covalently attached to a nanoreporter. The constant region can be attached at either the 5' region or the 3' region of a nanoreporter, and may be utilized for capture and immobilization of a nanoreporter for imaging or detection, such as by attaching to a solid substrate a sequence that is complementary to the constant region. In certain aspects, the constant region contains 2, 3, 4, 5, 6, 7, 8, 9, 10, or more tandemly-repeated sequences, wherein the repeat sequences each comprise about 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more nucleotides, including about 12-18, 13-17, or about 14-16 nucleotides.

[0044] The nanoreporters described herein comprise synthetic, designed sequences. In some embodiments, the nanoreporters described herein provide improved consistency in the label

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1117

brightness. Consistent brightness as described herein refers to the strength, size and/or the intensity of the signal produced by the labeled segments. That is, in some embodiments, the strength, size and/or intensity of the signal of the labeled segments will be similar among the segments that have the same label attached thereto. As described above, this improved consistency leads to more robust data. In some embodiments, a nucleotide is spaced at least an average of 8, 9, 10, 12, 15, 16, 20, 30, or 50 bases apart. In some embodiments, a nucleotide is spaced at least an average of 8 to 16 bases apart. In some embodiments, a nucleotide is spaced at least an average of 8 bases apart. In some embodiments, the sequences contain a fairly regularly-spaced pattern of a nucleotide (e.g. adenine) residue in the backbone. This allows for a regularly spaced complementary nucleotide in the complementary polynucleotide sequence having attached thereto a detectable molecule. For example, in some embodiments, when the nanoreporter sequences contain a fairly regularly-spaced pattern of adenine residues in the backbone, whose complement is a regularly-spaced pattern of uridine (U) residues in complementary RNA segments, the in vitro transcription of the segments can be done using an aminoallyl-modified uridine base, which allows the covalent amine coupling of dye molecules at regular intervals along the segment. In some embodiments, the sequences contain about the same number or percentage of a nucleotide (e.g. adenine) that is spaced at least an average of 8, 9, 10, 12, 15, 16, 20, 30, or 50 bases apart in the sequences. This allows for similar number or percentages in the complementary polynucleotide sequence having attached thereto a detectable molecule. Thus, in some embodiments, the sequences contain a nucleotides that is not regularly-spaced but that is spaced at least an average of 8, 9, 10, 12, 15, 16, 20, 30, or 50 bases apart, the number of nucleotides in the sequences can vary depending of the desired brightness for the nanoreporter. In some embodiments, 20%, 30%, 50%, 60%, 70%, 80%, 90% or 100% of the complementary nucleotide is coupled to a detectable molecule. For instance, in some embodiments, when the nanoreporter sequences contain a similar percentage of adenine residues in the backbone and the in vitro transcription of the complementary segments is done using an aminoallyl-modified uridine base, 20%, 30%, 50%, 60%, 70%, 80%, 90% or 100% of the aminoallyl-modified uridine base can be coupled to a detectable molecule. Alternatively, the ratio of aminoallyl-modified uridine bases and uridine bases can be changed during the in vitro transcription process to achieve the desired brightness. For example, in vitro transcription process can take place in the presence of a mixture with a ratio of 1/1 of uridine to aminoallyl-modified uridine bases, when some or all the aminoallyl-modified uridine bases can be coupled to a detectable molecule. Thus, a person of ordinary skill in the art will understand that there are several methods in which consistent brightness among the nanoreporters can be achieved.

[0045] In some embodiments, the nanoreporters described herein have a fairly consistent melting temperature (Tm). In some embodiments, the Tm of the complementary polynucleotide sequences when hybridized to its label attachment regions in the nanoreporters described herein are higher than the Tm of a polynucleotide sequences complementary to a M13 DNA template when hybridized to a

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1118

nanoreporter probe comprising the M13 DNA. Without intending to be limited to any theory, the Tm of the nanoreporters described herein provides for stronger bonds between the nanoreporter backbone and the complementary polynucleotide sequence having attached thereto a detectable molecule, therefore, preventing dissociation during synthesis and hybridization procedures. In addition, the consistent Tm among a population of nanoreporters allows for the synthesis and hybridization procedures to be tightly optimized, as the optimal conditions are the same for all spots and positions. In some embodiments, the sequences of the nanoreporters have a 50% guanine/cytosine (G/C), with no more than three G's in a row. Thus, in some embodiments, the invention provides a population of nanoreporters in which the Tm among the nanoreporters in the population is fairly consistent. In some embodiments, the invention provides a population of nanoreporters in which the Tm of the complementary polynucleotide sequences when hybridized to its label attachment regions is about 80 ^{0}C, 85^{0}C, 90^{0}C, 100^{0}C or higher. In some embodiments, the invention provides a population of nanoreporters in which the Tm of the complementary polynucleotide sequences when hybridized to its label attachment regions is about 80 ^{0}C or higher.

[0046] In some embodiments, the nanoreporters described herein have minimal or no secondary structures, such as any stable intra-molecular base-paring interaction (e.g. hairpins). Without intending to be limited to any theory, the minimal secondary structure in the nanoreporters provides for better hybridization between the nanoreporter backbone and the polynucleotide sequence having attached thereto a detectable molecule. In addition, the minimal secondary structure in the nanoreporters provides for better detection of the detectable molecules in the nanoreporters. In some embodiments, the nanoreporters described herein have no significant intra-molecular pairing under annealing conditions of 75^{0}C, 1XSSPE. Secondary structures can be predicted by programs known in the art such as MFOLD. In some embodiments, the nanoreporters described herein contain less than 1% of inverted repeats in each strand, wherein the inverted repeats are 9 bases or greater. In some embodiments, the nanoreporters described herein contain no inverted repeats in each strand. In some embodiments, the nanoreporters do not contain any inverted repeat of 9 nucleotides or greater across a sequence that is 1100 base pairs in length. In some embodiments, the nanoreporters do not contain any inverted repeat of 7 nucleotides or greater across any 100 base pair region. In some embodiments, the nanoreporters described herein contain less than 1% of inverted repeats in each strand, wherein the inverted repeats are 9 nucleotides or greater across a sequence that 1100 base pairs in length. In some embodiments, the nanoreporters described herein contain less than 1% of inverted repeats in each strand, wherein the inverted repeats are 7 nucleotides or greater across any 100 base pair region. In some embodiments, the nanoreporters described herein contain a skewed strand specific content such that one strand is CT-rich and the other is GA rich.

[0047] The invention provides unique nanoreporters. In some embodiments, the nanoreporters described herein contain less that 1% of direct repeats. In some embodiments, the nanoreporters described herein contain no direct repeats. In some embodiments, the nanoreporters do not contain

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1119

any direct repeat of 9 nucleotides or greater across a sequence that 1100 base pairs in length. In some embodiments, the labeled nanoreporters do not contain any direct repeat of 7 nucleotides or greater across any 100 base pair region. In some embodiments, the nanoreporters described herein contain less than 1% of direct repeats in each strand, wherein the direct repeats are 9 nucleotides or greater across a sequence that 1100 base pairs in length. In some embodiments, the nanoreporters described herein contain less than 1% of direct repeats in each strand, wherein the direct repeats are 7 nucleotides or greater across any 100 base pair region. In some embodiments, the nanoreporters described herein contain less than 85, 80, 70, 60, 50, 40, 30, 20, 10, or 5 % homology to any other sequence used in the backbones or to any sequence described in the REFSEQ public database. In some embodiments, the nanoreporters described herein contain less than 85, % homology to any other sequence used in the backbones or to any sequence described in the REFSEQ public database. In some embodiments, the nanoreporters described herein contain less than 20, 16, 15, 10, 9, 7, 5, 3, 2 contiguous bases of homology to any other sequence used in the backbones or to any sequence described in the REFSEQ public database. In some embodiments, the nanoreporters described herein have no more than 15 contiguous bases of homology and no more than 85% identity across the entire length of the nanoreporter to any other sequence used in the backbones or to any sequence described in the REFSEQ public database.

[0048] In some embodiments, the sequence characteristics of the nanoreporter probes described herein provide improved detection of a target molecule. For instance, the binding of the nanoreporter probes to target molecules which results in the identification of the target molecules can be performed by individually detecting the presence of the nanoreporter. This can be performed by individually counting the presence of one or more of the nanoreporter molecules in a sample. In some embodiments where such counting methods are used, the nanoreporter probes described herein allow for an increased in the number of counts. In some embodiments, the number of molecular counts above background of said molecular complex after normalization of the sample is higher than 300, 400, 450, 500, 600, 700, 800, 900 or 1000 molecular counts. In some embodiments, the number of molecular counts above background of said molecular complex after normalization of the sample is higher than 400 molecular counts. In some embodiments, the percentage of valid molecular counts of the nanoreporter probes described herein is higher that about 10, 11, 12, 12.5, 14, 15, 20, 25, 30, 35, 40, 50, 60, 70, 80 or 90 %. In some embodiments, the percentage of valid molecular counts of the nanoreporter probes described herein is higher that about 10 %. In some embodiments, the percentage of valid molecular counts of the nanoreporter probes described herein is higher that about 12.5 %. In some embodiments, the number of molecular counts above background of the nanoreporter described herein after normalization of said sample is at least 2, 3, 5, 6, 10, 20, 30, 40, 50, 60, 70, 80, 90 or 100 fold higher than a comparable nanoreporter probe comprising M13 DNA. A comparable nanoreporter comprising M13 DNA is a nanoreporter comprising the same target specific region attached to a M13 DNA backbone. Examples of comparable nanoreporter probes comprising M13

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1120

DNA are described in the Examples section. In some embodiments, the number of molecular counts above background of the nanoreporter described herein after normalization of said sample is at least 2 fold higher than a comparable nanoreporter comprising M13 DNA. In some embodiments, the number of molecular counts above background of the nanoreporter described herein after normalization of said sample is at least 6 fold higher than a comparable nanoreporter comprising M13 DNA. In some embodiments, the number of molecular counts above background of the nanoreporter described herein after normalization of said sample is at least 20 fold higher than a comparable nanoreporter comprising M13 DNA. In some embodiments, the number of molecular counts above background of the nanoreporter described herein after normalization of said sample is at least 100 fold higher than a comparable nanoreporter comprising M13 DNA.

[0049] The elements of a nanoreporter can be found in a single molecular entity (a "singular" nanoreporter), or two distinct molecular entities (a "dual" nanoreporter). Each molecular entity may be composed of one molecule or more than one molecule attached to one another by covalent or non-covalent means. In some embodiments, each component of a dual nanoreporter has a target-specific sequence that binds to a different site on the same target molecule. This allows for smaller nanoreporter components with more efficient kinetics of binding of the nanoreporter to the target molecule and better signal:noise ratios resulting from the greater binding specificity. When using a dual nanoreporter system one of the nanoreporter probes may be unlabeled. In some embodiments, the unlabeled nanoreporter probe may comprise a capture region. In some embodiments, the unlabeled nanoreporter probe may comprise a target-specific region and a backbone that may be single stranded. In some embodiments, the unlabeled nanoreporter probe may comprise a target-specific region and a backbone that may be double stranded.

[0050] The complementary polynucleotide sequences attached to a nanoreporter backbone serve to attach detectable molecules, or label monomers, to the nanoreporter backbone. The complementary polynucleotide sequences may be directly labeled, for example, by covalent incorporation of one or more detectable molecules into the complementary polynucleotide sequence. Alternatively, the complementary polynucleotide sequences may be indirectly labeled, such as by incorporation of biotin or other molecule capable of a specific ligand interaction into the complementary polynucleotide sequence. In such instances, the ligand (*e.g.*, streptavidin in the case of biotin incorporation into the complementary polynucleotide sequence) may be covalently attached to the detectable molecule. Where the detectable molecules attached to a label attachment region are not directly incorporated into the complementary polynucleotide sequence, this sequence serves as a bridge between the detectable molecule and the label attachment region, and may be referred to as a bridging molecule, *e.g.*, a bridging nucleic acid.

[0051] The nucleic-acid based nanoreporter and nanoreporter-target complexes of the present invention comprise nucleic acids, which may be affinity-purified or immobilized using a nucleic acid, such as an oligonucleotide, that is complementary to the constant region or the nanoreporter or target

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1121

nucleic acid. As noted above, in some embodiments the nanoreporters comprise at least one constant region, which may serve as an affinity tag for purification and/or for immobilization (for example to a solid surface). The constant region typically comprises two or more tandemly-repeated regions of repeat nucleotides, such as a series of 15-base repeats. In such exemplary embodiments, the nanoreporter, whether complexed to a target molecule or otherwise, can be purified or immobilized by an affinity reagent coated with a 15-base oligonucleotide which is the reverse complement of the repeat unit.

[0052] Nanoreporters, or nanoreporter-target molecule complexes, can be purified in two or more affinity selection steps. For example, in a dual nanoreporter, one probe can comprise a first affinity tag and the other probe can comprise a second (different) affinity tag. The probes are mixed with target molecules, and complexes comprising the two probes of the dual nanoreporter are separated from unbound materials (*e.g.*, the target or the individual probes of the nanoreporter) by affinity purification against one or both individual affinity tags. In the first step, the mixture can be bound to an affinity reagent for the first affinity tag, so that only probes comprising the first affinity tag and the desired complexes are purified. The bound materials are released from the first affinity reagent and optionally bound to an affinity reagent for the second affinity tag, allowing the separation of complexes from probes comprising the first affinity tag. At this point only full complexes would be bound. The complexes are finally released from the affinity reagent for the second affinity tag and then preferably stretched and imaged. The affinity reagent can be any solid surface coated with a binding partner for the affinity tag, such as a column, bead (*e.g.*, latex or magnetic bead) or slide coated with the binding partner. Immobilizing and stretching nanoreporters using affinity reagents is fully described in U.S. Provisional Application no. 60/753,816 by Sean M. Ferree and Dwayne L. Dunaway, entitled "Compositions Comprising Oriented, Immobilized Macromolecules and Methods for Their Preparation," filed on December 23, 2005, which is incorporated by reference herein in its entirety.

[0053] The sequence of signals provided by the label monomers associated with the various label attachment regions of the backbone of a given nanoreporter allows for the unique identification of the nanoreporter. For example, when using fluorescent labels, a nanoreporter having a unique identity or unique spectral signature is associated with a target-specific sequence that recognizes a specific target molecule or a portion thereof. When a nanoreporter is exposed to a mixture containing the target molecule under conditions that permit binding of the target-specific sequence(s) of the nanoreporter to the target molecule, the target-specific sequence(s) preferentially bind(s) to the target molecule. Detection of the nanoreporter signal, such as the spectral code of a fluorescently labeled nanoreporter, associated with the nanoreporter allows detection of the presence of the target molecule in the mixture (qualitative analysis). Counting all the label monomers associated with a given spectral code or signature allows the counting of all the molecules in the mixture associated with the target-specific sequence coupled to the nanoreporter (quantitative analysis). Nanoreporters are thus useful for the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1122

diagnosis or prognosis of different biological states (*e.g.*, disease vs. healthy) by quantitative analysis of known biological markers. Moreover, the exquisite sensitivity of single molecule detection and quantification provided by the nanoreporters of the invention allows for the identification of new diagnostic and prognostic markers, including those whose fluctuations among the different biological states is too slight detect a correlation with a particular biological state using traditional molecular methods. The sensitivity of nanoreporter-based molecular detection permits detailed pharmacokinetic analysis of therapeutic and diagnostic agents in small biological samples.

[0054] Many nanoreporters, referred to as singular nanoreporters, are composed of one molecular entity. However, to increase the specificity of a nanoreporter and/or to improve the kinetics of its binding to a target molecule, a nanoreporter can be a dual nanoreporter composed of two molecular entities, each containing a different target-specific sequence that binds to a different region of the same target molecule. In a dual nanoreporter, at least one of the two molecular entities is labeled. The other molecular entity need not necessarily be labeled. Such unlabeled components of dual nanoreporters may be used as capture probes (*see* Figures 1 and 2) and optionally have affinity tags attached, such as biotin, which are useful to immobilize and/or stretch the complex containing the dual nanoreporter and the target molecule to allow visualization and/or imaging of the complex. For instance, in some embodiments, a dual nanoreporter with a 6-position nanoreporter code, using one 6-position coded nanoreporter and a capture probe. In some embodiments a dual nanoreporter with a 7-position nanoreporter code can be used, using one 8-position nanoreporter component and one single-position nanoreporter component. In some embodiments, a dual nanoreporter with an 6-position nanoreporter code, using one capture probe with an affinity tag and one 6-position nanoreporter component. In some embodiment an affinity tag is optionally included and can be used to purify the nanoreporter or immobilize the nanoreporter (or nanoreporter-target molecule complex) for the purpose of imaging.

[0055] Because of their modular structures, nanoreporters may be assembled and labeled in a variety of different ways. For example, a nanoreporter backbone can be attached to a target-specific sequence (for example by hybridization and, optionally, ligation), and the structure comprising the backbone and target-specific sequence attached to one or more complementary polynucleotide sequences having attached thereto, either directly or indirectly, a detectable molecule. Alternatively, the nanoreporter backbone can first be attached to one or more complementary polynucleotide sequences, and the backbone structure then attached to a target specific sequence. Thus, unless stated otherwise, a discussion or listing of steps in nanoreporter assembly does not imply that a specific route of assembly must be followed.

[0056] Nanoreporters' syntheses can be performed by any suitable methods known in the art. Double-stranded plasmids carrying nanoreporter sequences (e.g. DV1 sequences) can be grown in some bacterial strains under conditions of low temperature (no greater than 34°C). Nanoreporter linear single-stranded backbones can be made from double-stranded plasmid DNA using a four-step

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1123

protocol that includes linearization with a restriction enzyme, dephosphorylation with a thermolabile phosphatase, digestion with a second restriction enzyme to separate the cloning vector from the backbone sequence, and a strand-specific lambda exonuclease digestion which leaves only one strand of the backbone fragment intact. Figure 9 shows an example of a vector that by used for nanoreporters' syntheses.

[0057] Nanoreporter assembly and use is exemplified herein largely by way of description of a variety of nucleic acid-based nanoreporters. Illustrative embodiments of partially and fully assembled nanoreporters are listed below.

[0058] At its simplest, the invention provides a nucleic acid backbone having a plurality (e.g. 3) of label attachment regions that are capable of being labeled and resolved, each of which is made of a rationally-designed nucleotide sequence. These sequences encompass one or more, or all the characteristics described herein that make the nanoreporter more stable. Examples of polynucleotide templates that may be utilized to generate these designed label attachment regions are set forth in the polynucleotide sequences of SEQ ID NOS:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, and 24. The label attachment regions of each individual backbone are arranged in a unique linear order or combination, making each individual backbone unique as compared to the other backbones in a population. Since the individual nanoreporters in a population are each made of a unique backbone, each nanoreporter is similarly unique as compared to the other nanoreporters in a population. In some embodiments, each label attachment region is unique when compared to the other label nanoreporter region in the backbone.

[0059] In some embodiments, the nucleotide sequences of the individual label attachment regions within each nanoreporter are different from the nucleotide sequences of the other label attachment regions within that nanoreporter, preventing rearrangements, such recombination, sharing or swapping of the label polynucleotide sequences, and thereby improving the molecular stability. The number of label attachment regions to be formed on a backbone is based on the length and nature of the backbone, the means of labeling the nanoreporter, as well as the type of label monomers providing a signal to be attached to the label attachment regions of the backbone. In some embodiments, the complementary nucleotide sequence of each label attachment region is assigned a specific detectable molecule.

[0060] The invention also provides labeled nanoreporters wherein one or more label attachment regions are attached to a corresponding detectable molecule, each detectable molecule providing a signal. For example, in some embodiments, a labeled nanoreporter according to the invention is obtained when at least three detectable molecules are attached to three corresponding label attachment regions of the backbone such that these labeled label attachment regions, or spots, are distinguishable based on their unique linear arrangement. A "spot," in the context of nanoreporter detection, is the aggregate signal detected from the label monomers attached to a single label attachment site on a nanoreporter, and which, depending on the size of the label attachment region and the nature (*e.g.*,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1124

primary emission wavelength) of the label monomer, may appear as a single point source of light when visualized under a microscope. Spots from a nanoreporter may be overlapping or non-overlapping. The nanoreporter code that identifies that target molecule can comprise any permutation of the length of a spot, its position relative to other spots, and/or the nature (*e.g.*, primary emission wavelength(s)) of its signal. Generally, for each probe or probe pair of the invention, adjacent label attachment regions are non-overlapping, and/or the spots from adjacent label attachment regions are spatially and/or spectrally distinguishable, at least under the detection conditions (*e.g.*, when the nanoreporter is immobilized, stretched and observed under a microscope, as described in U.S. Application Serial No. 61/029,220, incorporated herein by reference).

[0061] Occasionally, reference is made to a spot size as a certain number of bases or nucleotides. As would be readily understood by one of skill in the art, this refers to the number of bases or nucleotides in the corresponding label attachment region.

[0062] The order and nature (*e.g.*, primary emission wavelength(s), optionally also length) of spots from a nanoreporter serve as a nanoreporter code that identifies the target molecule capable of being bound by the nanoreporter through the nanoreporter's target specific sequence(s). When the nanoreporter is bound to a target molecule, the nanoreporter code also identifies the target molecule. Optionally, the length of a spot can be a component of the nanoreporter code.

[0063] Detectable molecules providing a signal associated with different label attachment regions of the backbone can provide signals that are indistinguishable under the detections conditions ("like" signals), or can provide signals that are distinguishable, at least under the detection conditions (*e.g.*, when the nanoreporter is immobilized, stretched and observed under a microscope).

[0064] The invention also provides a nanoreporter wherein two or more detectable molecules are attached to a label attachment region. The signal provided by the detectable molecules associated with said label attachment region produces an aggregate signal that is detected. The aggregate signal produced may be made up of like signals or made up of at least two distinguishable signals (*e.g.*, spectrally distinguishable signals).

[0065] In one embodiment, the invention provides a nanoreporter wherein at least three detectable molecules providing like signals are attached to three corresponding label attachment regions of the backbone and said three detectable molecules are spatially distinguishable. In another embodiment, the invention provides a nanoreporter wherein at least three detectable molecules providing three distinguishable signals are attached to three neighboring label attachment regions, for example three adjacent label attachment regions, whereby said at least three label monomers are spectrally distinguishable.

[0066] The invention also provides a nanoreporter wherein the spots providing like or unlike signals are separated by a spacer region, whereby interposing the spacer region allows the generation of dark spots, which expand the possible combination of uniquely detectable signals. The term "dark spot" refers to a lack of signal from a label attachment site on a nanoreporter. Dark spots can be

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1125

incorporated into the nanoreporter code to add more coding permutations and generate greater nanoreporter diversity in a nanoreporter population. In one embodiment, the spacer regions have a length determined by the resolution of an instrument employed in detecting the nanoreporter.

[0067] The invention provides a nanoreporter with one or more "double spots." Each double spot contains two or more (*e.g.*, three, four or five) adjacent spots that provide like signals without being separated by a spacer region. Double spots can be identified by their sizes.

[0068] A detectable molecule providing a signal according to the invention may be attached covalently or non-covalently (*e.g.*, via hybridization) to a complementary polynucleotide sequence that is attached to the label attachment region. The label monomers may also be attached indirectly to the complementary polynucleotide sequence, such as by being covalently attached to a ligand molecule (*e.g.*, streptavidin) that is attached through its interaction with a molecule incorporated into the complementary polynucleotide sequence (*e.g.*, biotin incorporated into the complementary polynucleotide sequence), which is in turn attached via hybridization to the backbone.

[0069] Certain aspects of the invention also provide a nanoreporter associated with a uniquely detectable signal, such as a spectral code, determined by the sequence of signals provided by the label monomers attached (*e.g.*, indirectly) to label attachment regions on the backbone of the nanoreporter, whereby detection of the signal allows identification of the nanoreporter.

[0070] In one embodiment, the invention provides a nanoreporter further comprising an affinity tag attached to the nanoreporter backbone, such that attachment of the affinity tag to a support allows backbone stretching and resolution of signals provided by label monomers corresponding to different label attachment regions on the backbone. Nanoreporter stretching may involve any stretching means known in the art including but not limited to, means involving physical, hydrodynamic or electrical means. The affinity tag may comprise a constant region.

[0071] A nanoreporter according to the invention can further include a target-specific sequence coupled to the backbone. The target-specific sequence is selected to allow the nanoreporter to recognize, bind or attach to a target molecule. The nanoreporters of the invention are suitable for identification of target molecules of all types. For example, appropriate target-specific sequences can be coupled to the backbone of the nanoreporter to allow detection of a target molecule. Preferably the target molecule is DNA (including cDNA), RNA (including mRNA and cRNA), a peptide, a polypeptide, or a protein.

[0072] One embodiment of the invention provides increased flexibility in target molecule detection with label monomers according to the invention. In this embodiment, a dual nanoreporter comprising two different molecular entities, each with a separate target-specific region, at least one of which is labeled, bind to the same target molecule. Thus, the target-specific sequences of the two components of the dual nanoreporter bind to different portions of a selected target molecule, whereby detection of the spectral code associated with the dual nanoreporter provides detection of the selected target molecule in a biomolecular sample contacted with said dual nanoreporter.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1126

[0073] The invention also provides a method of detecting the presence of a specific target molecule in a biomolecular sample comprising: (i) contacting said sample with a nanoreporter as described herein (*e.g.*, a singular or dual nanoreporter) under conditions that allow binding of the target-specific sequences in the dual nanoreporter to the target molecule and (ii) detecting the spectral code associated with the dual nanoreporter. Depending on the nanoreporter architecture, the dual nanoreporter may be labeled before or after binding to the target molecule.

[0074] The uniqueness of each nanoreporter probe in a population of probe allows for the mutiplexed analysis of a plurality of target molecules. For example, in some embodiments, each nanoreporter probe contains six label attachment regions, where each label attachment region of each backbone is different from the other label attachment regions in that same backbone. If the label attachment regions are going to be labeled with one of four colors and there are 24 possible unique sequences for the label attachment regions and each label attachment region is assigned a specific color, each label attachment region in each backbone will consist of one of four sequences. There will be 4096 possible nanoreporters in this example. The number of possible nanoreporters can be increased, for example, by increasing the number of colors, increasing the number of unique sequences for the label attachment regions and/or increasing the number of label attachment regions per backbone. Likewise the number of possible nanoreporters can be decreased by decreasing the number of colors, decreasing the number of unique sequences for the label attachment regions and/or decreasing the number of label attachment regions per backbone.

[0075] In certain embodiments, the methods of detection are performed in multiplex assays, whereby a plurality of target molecules are detected in the same assay (a single reaction mixture). In a preferred embodiment, the assay is a hybridization assay in which the plurality of target molecules are detected simultaneously. In certain embodiments, the plurality of target molecules detected in the same assay is, at least 2, at least 5 different target molecules, at least 10 different target molecules, at least 20 different target molecules, at least 50 different target molecules, at least 75 different target molecules, at least 100 different target molecules, at least 200 different target molecules, at least 500 different target molecules, or at least 750 different target molecules, or at least 1000 different target molecules. In other embodiments, the plurality of target molecules detected in the same assay is up to 50 different target molecules, up to 100 different target molecules, up to 150 different target molecules, up to 200 different target molecules, up to 300 different target molecules, up to 500 different target molecules, up to 750 different target molecules, up to 1000 different target molecules, up to 2000 different target molecules, or up to 5000 different target molecules. In yet other embodiments, the plurality of target molecules detected is any range in between the foregoing numbers of different target molecules, such as, but not limited to, from 20 to 50 different target molecules, from 50 to 200 different target molecules, from 100 to 1000 different target molecules, from 500 to 5000 different target molecules, and so on and so forth.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1127

[0076] In certain embodiments, the invention is directed to detecting different splice forms of the same RNA. The different splice forms can be detected using a plurality of nanoreporter probes, each with a different target-specific sequence complementary to a different exon of the same gene.

[0077] In addition to the qualitative analytical capabilities provided by the nanoreporters of the invention and the analytical techniques based thereon, the nanoreporters of the invention are uniquely suitable for conducting quantitative analyses. By providing a one to one binding between the nanoreporters (whether singular or dual nanoreporters) of the invention and their target molecules in a biomolecular sample, all or a representative portion of the target molecules present in the sample can be identified and counted. This individual counting of the various molecular species provides an accurate and direct method for determining the absolute or relative concentration of the target molecule in the biomolecular sample. Moreover, the ability to address each molecule in a mixture individually leverages benefits of miniaturization including high sensitivity, minimal sample quantity requirements, high reaction rates which are afforded by solution phase kinetics in a small volume, and ultimately very low reagent costs.

[0078] As will be appreciated from the description and examples provided below, the present invention provides numerous advantages. For example, the complex modularity in forming nanoreporters according to the invention allows for systematic creation of libraries of unique nanoreporters having a very high degree of diversity (*e.g.*, millions of uniquely recognizable nanoreporters). This modularity allows flexibility in customizing nanoreporter populations to specific applications which in turn provides significant manufacturing efficiencies. Another advantage that will be appreciated through the following description stems from the flexibility in assembling the nanoreporters of the invention. That is, due to their modular structure, the nanoreporters of the invention can be assembled prior to shipment to a point of use or assembled at the point of use.

Dual Nanoreporters

[0079] Figures 1 and 2 illustrate certain embodiments of the invention involving dual nanoreporters. In some embodiments, each of the two components of the nanoreporter is labeled, such that the nanoreporter's spectral code is formed only when the two components of the nanoreporter come together upon binding of the dual nanoreporter to its target molecule. However, in a dual nanoreporter, it is not necessary that both components are labeled. For example, as depicted in Figures 1 and 2, one component of a dual nanoreporter is labeled with the nanoreporter code, and the other component attached to an affinity tag (arrow) that is useful to immobilize the nanoreporter for stretching and visualization.

Definitions

[0080] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by those of ordinary skill in the art to which the invention belongs.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1128

Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, preferred methods and materials are described. For the purposes of the present invention, the following terms are defined below.

[0081] The articles "a" and "an" are used herein to refer to one or to more than one (*e.g.* to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element.

[0082] The terms "complementary" with respect to polynucleotides refer to polynucleotides related by the base-pairing rules. For example, the sequence "A-G-T," is complementary to the sequence "T-C-A." Complementarity may be partial, in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be complete or total complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands.

[0083] The terms “polynucleotides”, “nucleic acids”, “nucleotides” and “oligonucleotides” are used interchangeably. They refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. The following are non-limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component.

[0084] The recitations "sequence identity" or, for example, comprising a "sequence 50% identical to," as used herein, refer to the extent that sequences are identical on a nucleotide-by-nucleotide basis over a window of comparison. Thus, a "percentage of sequence identity" may be calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (*e.g.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity.

[0085] Terms used to describe sequence relationships between two or more polynucleotides include "homology", "comparison window", "sequence identity", "percentage of sequence identity" and "substantial identity". Because two polynucleotides may each comprise (1) sequence (*e.g.*, only a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) sequence that is divergent between the two polynucleotides, sequence comparisons between two

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1129

(or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity. A "comparison window" refers to a conceptual segment of at least 6 contiguous positions, usually about 50 to about 100, more usually about 100 to about 150 in which a sequence is compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. The comparison window may comprise additions or deletions (*i.e.*, gaps) of about 20% or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by computerized implementations of algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Drive Madison, WI, USA) or by inspection and the best alignment (*i.e.*, resulting in the highest percentage homology over the comparison window) generated by any of the various methods selected. Reference also may be made to the BLAST family of programs as for example disclosed by Altschul *et al.*, 1997, *Nucl. Acids Res.* 25:3389. A detailed discussion of sequence analysis can be found in Unit 19.3 of Ausubel *et al.*, "Current Protocols in Molecular Biology", John Wiley & Sons Inc, 1994-1998, Chapter 15.

Backbone Structure

[0086] The nanoreporter backbone of the present invention is a nucleic acid molecule, containing a plurality of label attachments regions (e.g. at least three label attachment regions) arranged in a linear combination to which label monomers can be directly or indirectly attached. In one embodiment, the nanoreporter backbone is a nucleic acid backbone in which the label attachment regions are single-stranded regions to which other nucleic acids, such as complementary oligonucleotide, complementary RNA sequences, or complementary DNA sequences, can attach by hybridization. In specific embodiments, the nanoreporter backbone is a single-stranded nucleic acid molecule.

[0087] In some embodiments, the backbone of the present invention is DNA. DNA based structures offer numerous advantages in the context of the present invention due at least in part to the vast universe of existing techniques and methodologies that allow manipulation of DNA constructs. As indicated above, the backbone may be single stranded.

[0088] Each backbone is comprised of a unique arrangement of label attachment regions. The label attachment regions of a nanoreporter backbone will vary in size depending on the method of labeling. In various embodiments, a label attachment region can have a length anywhere from 10 nm to 10,000 nm, but is more preferably from 50 nm to 5,000 nm, and is more preferably from 100 nm to 1,000 nm. In various embodiments, the label attachment region is from about 100 nm to about 500 nm, from about 150 nm to about 450 nm, from about 200 nm to about 400 nm, or from 250 to about 350 nm. In a preferred embodiment, the label attachment region corresponds closely to the size of a diffraction-limited spot, *e.g.*, the smallest spot that can be detected with standard optics, which is about 300 nm.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1130

[0089] In some embodiments, the label attachment regions of the present invention are made of polynucleotide sequences. For the nucleic acid backbones of the present invention, 1 nm corresponds to about 3 nucleotides; thus, an about 300 nm-label attachment region corresponds to a nucleotide sequence of about 900 bases. In certain embodiments, the label attachment region is from about 300 nucleotides to about 1.5kb, from about 450 nucleotides to about 1.35 kb, from about 0.6 kb to about 1.2 kb, or from 0.75 kb to about 1.05 kb. In certain aspects, the label attachment region is about 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1600, 1700, 1800, 1900, 2000, 2500, 3000, 3500, 4000, 4500, 5000 or more nucleotides, including all integers in between. In certain preferred embodiments, the label attachment region is a nucleotide sequence of about 1100 nucleotides in length.

[0090] A backbone can have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24-100 label attachment regions or more, including all integers in between. In certain embodiments, a backbone has 6 label attachment regions. In some embodiment the individual backbones have a set of distinct label attachment regions, each of which has a polynucleotide sequence that is different from the sequences of the other label attachment regions in that same backbone.

[0091] In some embodiments, a backbone has 6 unique label attachments regions. In some embodiments, each backbone comprises a plurality of different detectable molecules wherein said plurality of different detectable molecules in each backbone has a detectable signal that distinguishes it from other backbones in the population.

[0092] According to the present invention, each backbone in a population of backbones is distinguishable from the others by having a unique linear combination, order, or arrangement of label attachment regions. In addition, as one aspect of the present invention, the polynucleotide sequence of each label attachment region in a given backbone is different from the polynucleotide sequence of the other label attachment regions in that same backbone. In a linear arrangement of multiple label attachment regions on the same backbone, repeating more than one of the same label attachment regions can cause the plasmids from which the backbones are generated to be unstable during replication in *E. coli* and therefore difficult to propagate. In addition, the introduction of reverse-complement RNA segments to such a backbone may cause the formation of "knotted" reporters in which an RNA segment bridges and pulls together two identical label attachment regions, disrupting the linear sequence of the reporter and interfering with the accuracy and readability of the reporter code, among other problems. In embodiments in which plasmid instability and potentially poor reporter readability are a problem, these problems are avoided by constructing each individual backbone to have a set of distinct label attachment regions, each of which has a polynucleotide sequence that is different from the sequences of the other label attachment regions in that same backbone.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1131

[0093] In addition, as a further aspect of the present invention, these label attachment regions are selected from a population of unique, rationally-designed (*e.g.*, synthetic sequences) nucleotide sequences, each of which has a unique polynucleotide sequence represented by certain characteristics (*e.g.*, G/C content, G/C ratio, adenine repeats), as detailed herein, and each of which is designated a given detectable molecule. These rationally-designed sequences not only improve stability by preventing secondary structure formation (e.g. keeping linear sequence of the reporter to maintain accuracy and readability of the reporter code), but optimize the concentration and spacing of the detectable molecules attached to each label attachment region, thereby improving the readability of a given labeled backbone among a population of uniquely labeled backbones.

Label Attachment Regions

[0094] The present invention provides nanoreporter backbones that are synthetic nucleic acid molecules (DNA, RNA, or DNA/RNA hybrids), rationally-designed to have features that optimize labeling and detection of the nanoreporter, and that prevent secondary structure formation. In some embodiments of the invention, a nanoreporter backbone is a designed polynucleotide sequence comprising one or more sequences from 50 to 50,000 bases long.

[0095] In some embodiments, the nanoreporter backbone is designed to minimize predictable secondary structures. In some embodiments the nanoreporter backbone are devoid of any secondary structure. Putatative secondary structures (e.g. hairpins, folding, or internal base pairing) can be predicted by methods known in the art such as MFOLD. Without intending to be limited to any theory, predictable secondary structure in the nanoreporter structure can be minimized by avoiding inverted repeats and by skewing the backbone-specific content such that the backbone is CT or GA-rich. In some embodiments, the nanoreporter backbone does not have any significant intra-molecular paring under annealing conditions of 75°C, 1XSSPE. In some embodiments, the nanoreporter backbone has less than about 20, 15, 10, 5, 4, 3, 2, or 1% of inverted repeats, wherein the inverted repeats are 9 nucleotides or greater. In some embodiments, the backbone has less than about 1% of inverted repeats, wherein the inverted repeats are 9 nucleotides or greater. In some embodiments, the nanoreporters do not contain any inverted repeat of 9 nucleotides or greater across a sequence that 1100 base pairs in length. In some embodiments, the nanoreporters do not contain any inverted repeat of 7 nucleotides or greater across any 100 base pair region. In some embodiments, the nanoreporter backbone is designed to skew the strand-specific content of the backbone such that the backbone is CT or GA-rich. A CT or GA region is a region compose by any combination of the nucleotides C and T, or G and A. For instance, the smallest CT region is a region of two nucleotides wherein the two nucleotides are C and T. The regions can be 2, 3, 4, 5, 6, 7, 8, or more nucleotides in length. In some embodiments about 30, 35, 40, 45, 50, 55, 60, 65 or 70% of the nanoreporter backbone is compose of CT or GA regions. In some embodiments about 50 to 65 % of the nanoreporter backbone is composed of CT or GA regions. In some embodiments about 60 % of the nanoreporter backbone is

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1132

composed of CT or GA regions. In some embodiments, the nanoreporter backbone is designed to skew the strand-specific content of the backbone such that the backbone is CT -rich. In some embodiments about 60 % of the nanoreporter backbone is composed of CT regions.

[0096] In some embodiments, the nanoreporter backbone is designed to maximize the uniqueness of all of the sequence in a nanoreporter population by avoiding direct repeats and by screening out all sequence that had any significant homology to any other sequence used in other backbones in a population of nanoreporters or to any sequence described in the REFSEQ public database. In some embodiments, the nanoreporter backbone has less than about 20, 15, 10, 5, 4, 3, 2, or 1% of direct repeats. In some embodiments, the nanoreporter backbone has less than about 1% of direct repeats, wherein the direct repeats are 9 nucleotides or more in a sequence of 1100 base pair of length or 7 nucleotides or more across any 100 base pair region. In some embodiments, the nanoreporters described herein contain less than 85, 80, 70, 60, 50, 40, 30, 20, 10, or 5 % homology to any other sequence used in other backbones in a population of nanoreporters or to any sequence described in the REFSEQ public database. In some embodiments, the nanoreporters described herein contain less than 60 % homology to any other sequence used in other backbones in a population of nanoreporters or to any sequence described in the REFSEQ public database. In some embodiments, the nanoreporters described herein contain less than 50 % homology to any other sequence used in other backbones in a population of nanoreporters or to any sequence described in the REFSEQ public database. In some embodiments, the nanoreporter's backbone is designed to contain less than 85 % homology to any other sequence used in the backbones or to any sequence described in the REFSEQ public database. In some embodiments, the nanoreporters backbone is designed to contain less than 20, 16, 15, 10, 9, 7, 5, 3, 2 contiguous bases of homology to any other sequence used in the backbones or to any sequence described in the REFSEQ public database. In some embodiments, the nanoreporters backbone is designed to have no more than 15 contiguous bases of homology and no more than 85% identity across the entire length of the nanoreporter to any other sequence used in the backbones or to any sequence described in the REFSEQ public database.

[0097] The nucleic acid label attachment regions of the invention preferably do not have direct or inverted repeats that are greater than 12 bases in length. In other embodiments, the nucleic acid label attachment regions do not have direct or inverted repeats that are greater than about 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 or more nucleotides bases in length. In some embodiments, the nucleic acid label attachment regions do not contain any inverted repeat of 9 nucleotides or greater across a sequence that 1100 base pairs in length. In some embodiments, the nucleic acid label attachment regions do not contain any inverted repeat of 7 nucleotides or greater across any 100 base pair region.

[0098] Sequences may also be screened to avoid common six-base-cutter restriction enzyme recognition sites. Selected sequences may be additionally subjected to predicted secondary structure analysis, and those with the least secondary structure may be chosen for further evaluation. Any

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1133

program known in the art can be used to predict secondary structure, such as the MFOLD program (Zuker, 2003, *Nucleic Acids Res*. 31 (13):3406-15; Mathews *et al.*, 1999, *J. Mol. Biol*. 288:911-940).

[0099] The nanoreporter backbone, which in some embodiments is a single-stranded nucleic acid molecule, is designed to have one or more label attachment regions, comprising a regular pattern of a particular base (the "regularly-repeated base"). In such regions, the regularly-repeated base typically occurs with a minimal periodicity of about every nth residue, where n is any number, and preferably from 4 to 25. Preferably, not more than 25% of the regularly-repeated base in a label attachment region appears at other than said regular intervals. For example, if in a label attachment region of 100 nucleotides there are 12 thymine bases, and thymine is the regularly-repeated base, in this aspect of the invention not more than 25% of these, *i.e.*, 3 thymine bases, appear outside the regular pattern of thymines. Similarly, if in a label attachment region of 100 nucleotides there are 12 adenine bases, and adenine is the regularly-repeated base, in this aspect of the invention not more than 25% of these, *i.e.*, 3 adenine bases, appear outside the regular pattern of adenines. In specific embodiments, not more than 20%, not more than 15%, not more than 10%, not more than 9%, not more than 8%, not more than 7%, not more than 6%, not more than 5%, not more than 4%, not more than 3%, not more than 2% or not more than 1% of said base appears at other than said regular intervals in said region. In some embodiments, the bases are not regularly-spaced but the bases are at least an average of 8, 9, 10, 12, 15, 16, 18, 20, 30, or 50 nucleotides apart. In some embodiments, the bases are not regularly-spaced but the bases are at least an average of 8 nucleotides apart.

[00100] In certain embodiments, the regularly-repeated base in a given label attachment region is spaced at least about every 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or more nucleotide bases. In certain aspects, the regularly repeated base is spaced at about every 5 to 15 nucleotide bases, about every 6 to 12 nucleotide bases, about every 7 to 10 nucleotide bases, or about every 8 to 10 nucleotide bases. In certain aspects, the regularly repeated base is spaced at about every 8 nucleotide bases. In certain specific embodiments, an adenine base is repeated at about every 8 nucleotide bases. In certain specific embodiments, an adenine base is repeated at about every 15 to 20 nucleotide bases. In certain specific embodiments, an adenine base is repeated at about every 16 nucleotide bases.

[00101] The regularly repeated base may be present in a the polynucleotide sequence of a given label attachment region at about 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20% nucleotide content. The presence of a regularly repeated base at an average of about every 8 to 16 nucleotides in a label attachment region represents about 12.5% nucleotide content. In certain embodiments, the nucleotide content of the regularly repeated base may be about 12.5%. In certain aspects, the label attachment region comprises a selected number or selected percentage of regularly-repeated bases per region, which are distributed randomly except that they must be at least a certain minimum distance apart, such at least about 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or more bases apart.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1134

[00102] The regularly repeated base can be any nucleotide base of adenine, thymine, guanine, cytosine, or uracil. In certain embodiments, the regularly repeated base in the single-stranded label attachment region is adenine.

[00103] In certain aspects, the polynucleotide sequence of the label attachment regions is also designed to have a certain guanine/cytosine (G/C) content. For example, certain label attachment regions are designed to have an overall G/C content of about 45%, 46%, 47%, 48%, 49%, 50%, 51 %, 52%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, more preferably about 50%, as this provides a fairly consistent T_M among the numerous, unique polynucleotide sequences of the label attachment regions. In certain aspects, the polynucleotide sequences have no more than three guanines in a row, to minimize synthesis issues. The overall GC-content is also preferably consistent over relatively short stretches to make local Tm's similar. Without intending to be limited to any theory, when the nanoreporter is a double stranded nucleic acid the Tm of the nanoreporters described herein provides for stronger bonds between the nanoreporter backbone and the complementary polynucleotide sequence having attached thereto a detectable molecule, therefore, preventing dissociation during synthesis and hybridization procedures. In addition, the consistent Tm among the population of nanoreporters allows for the synthesis and hybridization procedures to be tightly optimized, as the optimal conditions are the same for all spots and positions. In some embodiments, the sequences of the nanoreporters have about 50% guanine/cytosine (G/C), with no more than three Gs in a row.

[00104] In certain aspects, the polynucleotide sequence of the label attachment regions is also designed to have a certain G/C ratio in a given strand. For example, certain embodiments the polynucleotide sequence of the label attachment regions are designed such that the complementary polynucleotide sequences have a G/C ratio of about 1/1, 2/1, 3/2, about 5/3, or about 7/5. In certain aspects, the G/C ratio is at least 1/1. In certain aspects, the G/C ratio is about 1/1, 2/1, 3/2, 5/3, or about 7/5 in a given strand. In certain aspects, the G/C ratio is at least 1/1 in the complementary polynucleotide sequence. In certain aspects, the G/C ratio is about 3/2 in the complementary polynucleotide sequence.

[00105] In certain specific embodiments, in which adenine is the regularly repeated base, spaced at least about every 8 nucleotides (about 12.5% adenine content), and the G/C content is about 50%, the thymine content may be about 37-38%, or about 37.5%. In certain specific embodiments, in which adenine is the regularly repeated base, and the G/C content is about 50% In some embodiments, the thymine content may be about 35-45%. In an exemplary label attachment region in which the regularly-repeated nucleotide is adenine and a GC content of about 50%, excess adenines may be utilized to make up the loss in abundance of T's. To generate the selected sequence, random sequences with fixed patterns of As ranging from every 4th base to every 25th base may be created and screened to minimize the presence of inverted and direct repeats.

[00106] In some embodiments, the present invention provides nanoreporter or nanoreporter label unit populations, for example nanoreporter or nanoreporter label unit libraries, that contain at least 3, at

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1135

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 45, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 200, at least 300, at least 400, at least 500, at least 750, or at least 1,000 unique nanoreporters, unique nanoreporter label units, or unique labeled nanoreporters. As used herein, "unique" when used in reference to a nanoreporter or nanoreporter label units within a population is intended to mean a nanoreporter that has a code that distinguishes it from other nanoreporters in the same population, e.g., each nanoreporter has a detectable signal that distinguishes it from other nanoreporters in said population. Typically, a nanoreporter is rendered "unique" by the particular linear combination of label attachment regions each having attached thereto a selected label monomer.

[00107] In specific embodiments, the present invention provides nanoreporter populations with at least 1,000, 5,000, at least 10,000, at least 20,000 or at least 50,000 unique nanoreporters or nanoreporter label units.

[00108] The nanoreporters in a population of nanoreporters can be singular nanoreporters, dual nanoreporters, or a combination thereof. The nanoreporters can be labeled or unlabeled. The population of nanoreporters described herein can comprise two or more nanoreporters.

[00109] In some embodiments, the invention provides for a population of nanoreporter backbones in which a base (e.g. adenine) is regularly spaced, the nanoreporters do not have any significant formation of secondary structures and the Tm among the nanoreporters in the population is fairly consistent. In some embodiments, the invention provides a population of nanoreporters in which the Tm of the complementary polynucleotide sequences when hybridized to its label attachment regions is about 80 ^{0}C or higher.

[00110] In some embodiments, the invention provides for a population of nanoreporter backbones wherein each nanoreporter backbone comprises a region comprising a plurality of label attachment regions covalently attached together in a linear combination, wherein each label attachment region comprises about 800 to 1300 nucleotide bases and has a G/C content of about 50%, wherein each selected label attachment region is different from the other selected label attachment regions, and wherein each label attachment region is hybridized to a complementary polynucleotide sequence having attached thereto one or more detectable molecules, wherein the complementary polynucleotide sequence has a G/C ratio of at least 1/1. In some embodiments, the complementary polynucleotide sequence has a G/C ratio of about 3/2. In some embodiments, the nanoreporter backbone is designed to skew the strand-specific content of the backbone such that the backbone is CT -rich. In some embodiments about 60 % of the nanoreporter backbone is composed of CT regions.

[00111] In some embodiments, the invention provides for a population of nanoreporter backbones wherein each nanoreporter backbone comprises a single-stranded nucleic acid backbone, said backbone comprising a plurality of label attachment regions covalently attached together in an unique linear combination, wherein said nanoreporter backbone comprises a base (e.g. adenine) regularly spaced, the nanoreporters backbones do not have any significant formation of secondary structures

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1136

and contain less than 85 % homology to any other sequence used in other backbones in the population or to any sequence described in the REFSEQ public database.

[00112] In some embodiments, the invention provides for a population of nanoreporter backbones wherein each nanoreporter backbone comprises a single-stranded nucleic acid backbone, said backbone comprising a plurality of label attachment regions covalently attached together in an unique linear combination, wherein said nanoreporter backbone has a G/C content of about 50%, the nanoreporters backbones do not have any significant formation of secondary structures and contain less than 85 % homology to any other sequence used in other backbones in the population or to any sequence described in the REFSEQ public database.

[00113] In some embodiments, the invention provides for a population of nanoreporter backbones wherein each nanoreporter backbone comprises a single-stranded nucleic acid backbone, said backbone comprising a plurality of label attachment regions covalently attached together in an unique linear combination, wherein said nanoreporter backbone has a G/C content of about 50%, the nanoreporters backbones do not have any significant formation of secondary structures and contain less than 1 % direct repeat sequences.

[00114] In some embodiments, the invention provides for a population of nanoreporter backbones wherein each nanoreporter backbone comprises a single-stranded nucleic acid backbone, said backbone comprising a plurality of label attachment regions covalently attached together in an unique linear combination, wherein the label attachment regions are selected from a population of unique, designed polynucleotide sequences, wherein each polynucleotide sequence is hybridized to an unique complementary polynucleotide sequence having attached thereto one or more detectable molecules, wherein each nanoreporter has a detectable signal that distinguishes it from other nanoreporters in said population, and wherein each nanoreporter is more stable than a comparable nanoreporter comprising M13 DNA.

[00115] In some embodiments, the invention provides for a population of nanoreporter backbones wherein each nanoreporter backbone comprises a single-stranded nucleic acid backbone, said backbone comprising a plurality of label attachment regions covalently attached together in an unique linear combination, wherein the label attachment regions are selected from a population of unique, designed polynucleotide sequences, wherein each polynucleotide sequence is hybridized to an unique complementary polynucleotide sequence having attached thereto one or more detectable molecules, wherein each nanoreporter has a detectable signal that distinguishes it from other nanoreporters in said population, wherein the Tm of the complementary polynucleotide sequences when hybridized to its label attachment regions in the nanoreporters described herein are higher than the Tm of a polynucleotide sequences complementary to a M13 DNA template when hybridized to a nanoreporter probe comprising the M13 DNA.

[00116] In some embodiments, the invention provides for a population of nanoreporter backbones wherein the sequence characteristics of the nanoreporter probes described herein provide improved

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1137

detection of a target molecule. For instance, when the binding of the nanoreporter probes to target molecules is performed by individually counting the presence of one or more molecules of the nanoreporter in a sample, the nanoreporter probes described herein allow for an increased in the number of counts. In some embodiments, the number of molecular counts above background of said molecular complex after normalization of the sample is higher than 300, 400, 450, 500, 600, 700, 800, 900 or 1000 molecular counts. In some embodiments, the nanoreporter probes described herein allow for an increased in the number of counts. In some embodiments, the number of molecular counts above background of said molecular complex after normalization of the sample is higher than 400 molecular counts. In some embodiments, the percentage of valid molecular counts of the nanoreporter probes described herein is higher that about 10, 11, 12, 12.5, 14, 15, 20, 25, 30, 35, 40, 50, 60, 70, 80 or 90 %. In some embodiments, the percentage of valid molecular counts of the nanoreporter probes described herein is higher that about 12.5 %. In some embodiments, the number of molecular counts above background of the nanoreporter probes described herein after normalization of said sample is at least 2, 5, 6, 10, 15, 20, 30, 40, 50, 60, 70, 80, or 90 fold higher than a comparable nanoreporter comprising M13 DNA. In some embodiments, the number of molecular counts above background of the nanoreporter probes described herein after normalization of said sample is at least 6 fold higher than a comparable nanoreporter comprising M13 DNA. In some embodiments, the number of molecular counts above background of the nanoreporter probes described herein after normalization of said sample is at least 2 fold higher than a comparable nanoreporter comprising M13 DNA.

[00117] A comparable M13 nanoreporter probe is a nanoreporter comprising the same target specific region as the nanoreporters probes described herein attached to a single–stranded backbone comprising a plurality of M13 DNA regions covalently attached together wherein each region is hybridized to a complementary polynucleotide sequence having attached thereto one or more detectable molecules.

[00118] The size of a nanoreporter probe population and the nature of the target-specific sequences of the nanoreporters within it will depend on the intended use of the nanoreporter. Nanoreporter probe populations can be made in which the target-specific sequences correspond to markers of a given cell type, including a diseased cell type. In a specific embodiment, the cell or tissue is a mammalian cell or tissue, and more preferably is a human cell or tissue.

[00119] In some embodiments, an appropriate sequence is divided into label attachment regions ranging from 50 bases to 2 kilobases long (could be longer). Label attachment regions can also range from about 200 bases to about 1800 bases, about 400 bases to about 1600 bases, about 600 bases to about 1500 bases, about 800 bases to about 1300 bases, about 900 bases to about 1200 bases, about 1000 bases to about 1150 bases, about 1050 bases to about 1150 bases, including combinations of these exemplary ranges and all integers in between (*e.g.*, 850, 950, etc). In certain embodiments the label attachment regions are about 1100 bases.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1138

[00120] In some embodiments, each label attachment region is a unique sequence, but contains a consistent and minimal number and spacing of adenines, or other selected base, in relation to the other label attachment regions in a given reporter sequence. These label attachment regions can interspersed with other regions whose sequence does not matter. The label attachment regions in a nanoreporter backbone can be of different lengths and/or have different regularly-repeated bases.

[00121] Specific examples of polynucleotide sequences that may be utilized as templates to synthesize single stranded label attachment regions include the polynucleotide sequences set forth in SEQ ID NOS:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, and 24. For these sequences, the recited regularly repeated base is thymine, which produces adenine as the regularly repeated based upon synthesis of the single stranded label attachment regions, typically in the form of a nanoreporter backbone. Also contemplated are variants of these polynucleotide sequences, including polynucleotides sequences having at least 60%, 70%, 80%, 90%, 95%, 98%, or more sequence identity to the polynucleotide sequences set forth in SEQ ID NOS:1-24, wherein such variants retain the desired or suitable characteristics of a label attachment region, as described in greater detail herein (*e.g.*, regularly-repeated base, G/C content, G/C ratio in a given strand, Tm, lack of inverted repeats). As it will be appreciated by one of ordinary skill in the art, other sequences can be designed to possess the desired or suitable characteristics of a label attachment region described herein.

[00122] An optimized start sequence for transcription by RNA polymerase T7, T3, or SP6 (beginning at position +1 of the transcript) can be added to the 5' end of each label attachment region. Restriction sites are optionally added at the boundaries of each label attachment region to allow specific addition or deletion of individual label attachment regions to the sequence using conventional cloning techniques.

[00123] The number of label attachment regions in a nanoreporter backbone preferably ranges from 1 to 50. In yet other embodiments, the number of synthetic label attachment regions in a nanoreporter backbone ranges from 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 rationally-designed label attachment regions to 15, 20, 30, 40, or 50 label attachment regions, or any range in between. In certain embodiments, the number of label attachment regions in a backbone is six.

[00124] In some embodiments, the regularly-repeated base in the label attachment region in a nanoreporter backbone, or its complementary regularly-repeated base in an annealed complementary polynucleotide sequence (or segment) can be used to attach label monomers, preferably light emitting label monomers, to the nanoreporter in a regular, evenly spaced pattern for better distribution of the nanoreporter signal. Preferably, where a label attachment region is labeled, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95% or at least 98% of occurrences of the regularly-repeated base is attached to at least one light-emitting label monomer, either by covalent attachment of a label monomer to a base, or by hybridization to a nucleic acid in which the complements of the regularly-repeated base are so-labeled.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1139

[00125] This percentage of occurrences can be measured by any means known in the art. In one method, the amount of nucleic acid produced in a labeling reaction is purified (for example, RNA can be purified using a Qiagen RNeasy kit) and subjected to UV spectrophotometry. The absorbance ("A") at the appropriate wavelengths is measured for each of the nucleic acid (260 nm) and the label monomer whose occurrence is to be measured (*e.g.*, 495 nm for Alexa Fluor 488; 590 nm for Alexa Fluor 594; 650 for Alexa Fluor 647; and 550 nm for Cy3). The absorbance of the nucleic acid is corrected by adjusting the value of the absorbance at 260 nm ("A260") to remove the "noise" contribution from the label monomer by subtracting the absorbance at the peak wavelength for the label monomer (A_{LM}) minus the correction factor for that label monomer. Where the nucleic acid is RNA, the number of label monomers per one thousand nucleotides is calculated according to the formula:

$$\frac{\text{no. of labelmonomers}}{\text{1000 nucleotides}} = \frac{\text{A260}}{\text{A}_{\text{LM}}} \times \frac{9010}{\text{EC}_{\text{LM}}} \times 1000$$

where EC_{LM} is the extinction coefficient for the label monomer. From this formula, the percentage of occurrences of the regularly-repeated base that are attached to a light-emitting label monomer can be calculated.

[00126] In some embodiments, the preferred regularly-repeating base in a label attachment region is adenine, so that the region can be labeled by hybridization to one or more complementary polynucleotide sequences (*e.g.*, RNA segments) in which the regularly-repeated base is uracil. In some embodiments, adenine is not regularly-repeated but is spaced at least an average of 8 bases apart. This permits the use of amino-allyl-modified UTPs, which are readily commercially available, as label monomer attachment sites, in an otherwise random sequence. Preferably, in addition to the regular periodicity of the label attachment regions, the regions (and the nucleic acid comprising them) contain minimal secondary structure and fairly consistent Tm.

[00127] The nucleic acids of the present invention can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the label attachment region and the annealed complementary polynucleotide sequences or segments, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the synthetic nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxyl methyl)uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethy1-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1140

methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl)uracil, (acp3)w, and 2,6-diaminopurine.

[00128] Alternatively, the synthetic nucleic acid can be produced biologically using a vector into which a nucleic acid has been subcloned. As one example, a linear single-stranded DNA backbone can be made from a double stranded plasmid DNA using a four step protocol that includes (i) linearization of the dsDNA with a restriction enzyme, (ii) dephosphorylation with a thermolabile phosphatase, (iii) digestion with a second restriction enzyme to separate the cloning vector from the backbone sequence, and (iv) digestion with a strand-specific lambda exonuclease digestion, leaving only one strand of the backbone fragment intact.

[00129] In various embodiments, the nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (*see* Hyrup *et al.*, 1996, *Bioorganic & Medicinal Chemistry* 40:5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in *Hyrup et al.,* 1996, *Bioorganic & Medicinal Chemistry* 4(1): 5-23; Perry-O'Keefe *et al.*, 1996, *Proc. Natl. Acad. Sci*. USA 93: 14670-675.

[00130] In an exemplary embodiment, the selected novel sequence can be constructed synthetically as double-stranded DNA by a commercial gene synthesis company and cloned in an oriented fashion into a "phagemid", a plasmid vector containing an M13 or f1 phage intergenic (IG) region which contains the cis-acting sequences necessary for DNA replication and phage encapsidation, such as pUC119. The appropriate orientation of the cloned insert relative to the phage origin of replication allows for the generation of a single-stranded DNA backbone which is the reverse complement of the RNA molecules generated by *in vitro* transcription for each label attachment region.

[00131] In order to generate the single-stranded DNA backbone of the novel reporter, the phagemid is transformed into an *E. coli* strain containing an F' episome. Subsequent infection of the transformed bacteria with a helper phage such as the M13 mutant K07 results in the secretion of the phagemid carrying the novel reporter sequence as a single-stranded, packaged phage from which the circular, single-stranded DNA is prepared using a standard protocol. This DNA is linearized and the vector portion is excised by annealing short, complementary oligonucleotides to either end of the novel reporter sequence to generate double-stranded restriction sites, followed by treatment with the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1141

appropriate restriction enzymes. Figure 9 shows an example of plasmid vector that may be used to clone, propagate, and generate the nanoreporters described herein.

Complementary Polynucleotide Sequences

[00132] In some embodiments, detectable molecules, or label monomers, that provide or emit signals which constitute all or part of the nanoreporter code are attached to label attachment region(s) of the nanoreporter backbone through a structure referred to herein as a complementary polynucleotide sequence. The detectable molecules can be directly (*e.g.*, covalently or noncovalently) attached to a complementary polynucleotide sequence, or indirectly attached to such a sequence (*e.g.*, through an intermediate component, such as a ligand).

[00133] In some embodiments, complementary polynucleotide sequences can be from 25 nucleotides to several kilobases (*e.g.*, 5kb) in length, and are preferably 50 nucleotides to 2kb in length. In specific embodiments, complementary polynucleotide sequences are about 25 to 250, 50 to 200, 50 to 150, or 50 to 100 nucleotides in length. In other embodiments, complementary polynucleotide sequences are about 500 to 2,000, 500 to 1,500, 500 to 1,000, 750 to 1,250, or 750 to 1,000 nucleotides in length. In certain aspects, complementary polynucleotide sequences are about 800, 850, 900, 950, 1100, 1150, or 1100 nucleotides in length, including all integers in between. Complementary polynucleotide sequences can be complementary RNA polynucleotide or complementary DNA polynucleotide. In preferred embodiment, complementary polynucleotide sequences are complementary RNA polynucleotides.

[00134] A detectable molecule or label monomer can be covalently attached to a complementary polynucleotide sequence before or after the sequence is attached to the label attachment region of a nanoreporter backbone. For example, in attaching a detectable molecule to a complementary polynucleotide sequence, the label can be covalently attached by incorporation of a nucleotide containing a detectable molecule into the nucleic acid during its synthesis, but before it is hybridized the label attachment region of the backbone. Alternatively, during the synthesis of a complementary polynucleotide sequence, a nucleotide containing a detectable molecule acceptor group can be included, and the detectable molecule added to the complementary polynucleotide sequence after its synthesis, either before or after it is hybridized to the label attachment region of the backbone. Alternatively, the label monomer can be indirectly attached to the complementary polynucleotide sequence, for example, by incorporating a nucleotide containing a ligand-binding molecule (*e.g.*, biotin) into the complementary polynucleotide sequence during synthesis, and by adding a ligand (*e.g.*, streptavidin) that is covalently attached to the detectable molecule.

[00135] In some embodiments, a complementary polynucleotide sequence can range anywhere from 20 nucleotides to more than 5 kb in length, depending on the method of assembly of the nanoreporter. For example, where a complementary polynucleotide sequence has covalently incorporated into it one or more label monomers that provide signals that are part of the nanoreporter code in the context of

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1142

the labeled nanoreporter, the complementary polynucleotide sequence is preferably about 100 to about 10,000 bases, more preferably 200 to about 2000 bases, and yet more preferably 700 to about 1200 nucleotides in length, or about 1100 base pairs in length, and is generally referred to herein as a segment, a dark segment being the complementary polynucleotide sequence prior to the incorporation of the label monomer (but, in a preferred embodiment, containing label monomer acceptor sites, such as amino allyl nucleotides), and a colored segment being one containing the desired label monomer or label monomers. The T_m of a segment when hybridized to its label attachment region preferably is >80°C, more preferably >90°C, in 825 mM Na^+ (5X SSC).

[00136] As with the label attachment regions, a complementary polynucleotide sequence may comprise a pattern of regularly-repeated bases or it may comprise a similar percentage of bases, such as a guanine, adenine, thymine, cytosine, or uracil repeat base. When the repeat base in the single-stranded label attachment region is adenine, the repeat base in the complementary polynucleotide sequence may be thymine if the complementary polynucleotide sequence is a complementary DNA polynucleotide, or uracil if the complementary polynucleotide sequence is a complementary RNA polynucleotide.

[00137] The regularly-repeated base in a given complementary polynucleotide sequence may be spaced about every 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or more nucleotide bases. In certain aspects, the regularly repeated base is spaced at about every 5 to 15 nucleotide bases, about every 6 to 12 nucleotide bases, about every 7 to 10 nucleotide bases, or about every 8 to 10 nucleotide bases. In certain aspects, the regularly repeated base is spaced about every 8 nucleotide bases. In certain aspects, the regularly repeated base is spaced about every 15 to 20 nucleotide bases. In certain aspects, the regularly repeated base is spaced about every 16 nucleotide bases.

[00138] In certain aspects, the regularly repeated base is amino-allyl modified. This modification allows ready attachment (*e.g.*, direct or indirect, as above) of a detectable molecule or label monomer to the regularly repeated base. The amino-allyl modified base may be a guanine, adenine, thymine, cytosine, or uracil base. In certain embodiments, the regularly repeated base is an amino-allyl modified uracil base.

[00139] In certain aspects, the detectable molecule or label monomer is attached to the regularly repeated base, such as via the amino-allyl modification of that base. In certain aspects, the detectable molecule is attached to all or almost all of the regularly repeated bases in a given polynucleotide complementary polynucleotide sequence. In certain embodiments, the detectable molecule is attached to about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% of the regularly repeated bases in a given complementary polynucleotide sequence.

[00140] In certain aspect, the detectable molecule or label monomer is attached to a base that is not regularly repeated but that is at least 8 bases apart such as via the amino-allyl modification of that

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1143

base. The brightness of the complementary polynucleotide sequence can vary according the number of bases in the sequence or the number of labeled bases in the sequence.

[00141] As one specific example, when adenine is regularly repeated in the single stranded label attachment region and is present at about every 8 nucleotide bases, an amino-ally modified uracil may also be present in the complementary RNA polynucleotide at about every 8 nucleotide bases. Also, the amino-ally modified uracils may incorporate randomly into the complementary polynucleotide sequences, such as by adjusting the ratio of uracil bases to modified uracil basis in the synthesis reaction. In these instances, the detectable molecule, such as a fluorescent dye, may be incorporated into all or almost all of the amino-allyl modified uracils (*e.g.*, at about every 8 bases) or it may be incorporated into the amino-allyl modified uracil bases at random.

[00142] By modifying the number of regularly repeated bases (*e.g.*, every 5, 6, 7, 8, 9, 10, or more bases), and incorporating the detectable molecule into every regularly repeated base, every other regularly repeated base, every third regularly repeated base, or at random, etc., or combinations thereof, the detectable molecule can be incorporated into about every 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more nucleotide bases, depending on the desired level of detectable signal.

[00143] In certain embodiments, the complementary polynucleotide sequences is an RNA polynucleotide comprising a sequence that is complementary to the single stranded label attachment regions generated by the polynucleotide template sequences set forth in SEQ ID NOS:1-24. A complementary RNA polynucleotide may be complementary to the full-length sequence of any of label attachment regions generated from the template sequences of SEQ ID NOS:1-24, or it may be complementary to a region within that polynucleotide sequence, such as a region comprising of 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or more nucleotide bases, including all integers in between.

[00144] In the embodiments where the complementary polynucleotide sequence is an RNA polynucleotide, the complementary RNA nucleotides (*e.g.*, segments) for each label attachment region may be made using methods known in the art, such as standard *in vitro* transcription (IVT) protocols. To synthesize the complementary RNA nucleotides, polymerase chain reaction ("PCR") primers may be designed to generate a double-stranded template beginning with an RNA polymerase promoter (T7, T3, or SP6) directly upstream (5') of the transcription start site and ending following the 3' restriction enzyme site. In certain aspects, the templates for these IVT reactions may include the synthetic DNA products of the templates described in SEQ ID NOS:1-24, or variant sequences thereof, which have been cloned downstream of T7, T3 or SP6 RNA polymerase promoters. Using such templates, *in vitro* transcription of RNA molecules may be performed in the presence of amino-allyl modified regularly-repeated base in the RNA (*e.g.*, UTP) and unmodified other bases (*e.g.*, ATP, CTP and GTP). This leads to an RNA product in which every regularly-repeated base (*e.g.*, U) is modified to allow covalent coupling of a label monomer at that position in the RNA molecule.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1144

[00145] The resulting amino-allyl modified transcripts (aa segments) may then be coupled to various fluorophores (*e.g.*, Alexa 488 (blue), Alexa 546 (green), Alexa 594 (yellow), and Alexa 647 (red) – other dyes can also be used). The Alexa fluorophores may be purchased as succinimidyl- or TFP-esters (Invitrogen™) and may be coupled to the transcripts via the formation of amide bonds using standard protocols. The level of dye-incorporation in a segment correlates with the number of possible attachment sites (amino allyl (aa) moieties). To make the brightest possible colored segments, 100% aaUTP is used. To change the brightness, a mixture of aaUTP and unmodified UTP in any ratio can be used in the IVT reactions to modify the number of aa sites present on an aa segment. The desired brightness varies with the optics of the detection system.

Detectable Molecules or Label Monomers

[00146] The nanoreporters of the present invention can be labeled with any of a variety of label monomers, such as a radioisotope, fluorochrome, dye, enzyme, nanoparticle, chemiluminescent marker, biotin, or other monomer known in the art that can be detected directly (*e.g.*, by light emission) or indirectly (*e.g.*, by binding of a fluorescently-labeled antibody). Generally, one or more of the label attachment regions in the nanoreporter is labeled with one or more label monomers, and the signals provided by the label monomers attached to the label attachment regions of a nanoreporter constitute a detectable code that identifies the target to which the target-specific region of the nanoreporter binds. In certain embodiments, the lack of a given signal from the label attachment region (*e.g.*, a dark spot) can also constitute part of the nanoreporter code.

[00147] In certain embodiments, such as when a nanoreporter backbone is constructed from a library (e.g. the exemplary label attachment regions templates of SEQ ID NOS:1-24), each unique label attachment region may be assigned a given detectable molecule. Merely by way of specific illustration, in a library of the 24 unique label attachment regions, e.g. products for templates SEQ ID NOS:1-24, the products of templates SEQ ID NOS:1, 5, 9, 13, 17, and 21 may be assigned a first detectable molecule, such as a blue fluorophore, the products of templates SEQ ID NOS:2, 6, 10, 14, 18, and 22 may be assigned a second detectable molecule, such as a green fluorophore, the products of templates SEQ ID NOS:3, 7, 11, 19, and 23 may be assigned a third detectable molecule, such as a yellow fluorophore, and the products of templates SEQ ID NOS: 4, 8, 12, 20, and 24 may be assigned a fourth detectable molecule, such as a red fluorophore. A person skilled in the art will appreciate that this example is merely illustrative, and that numerous combinations or alternatives may be utilized to accomplish the goal of creating a population of uniquely detectable nanoreporters, including the use of libraries of different sizes and having sequences other than those set forth in SEQ ID NOS:1-24.

[00148] Another example of label monomers that can be utilized by the invention are fluorophores. Several fluorophores can be used as label monomers for labeling nucleotides including, for example, fluorescein, tetramethylrhodamine, and Texas Red. Several different fluorophores are known, and more continue to be produced, that span the entire spectrum. Also, different formulations of the same

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1145

fluorophore have been produced for different applications. For example, fluorescein can be used in its isothiocynanate form (FITC), as mixed isomer or single isomer forms of carboxyfluorescein succinimidyl ester (FAM), or as isomeric dichlorotriazine forms of fluorescein (DTAF). These monomers are chemically distinct, but all emit light with a peak between 515-520 nm, thereby generating a similar signal. In addition to the chemical modifications of fluorescein, completely different fluorophores have been synthesized that have the same or very similar emission peaks as fluorescein. For example, the Oregon Green dye has virtually superimposable excitation and emission spectra compared to fluorescein. Other fluorophores such as Rhodol Green and Rhodamine Green are only slightly shifted in their emission peaks and so also serve functionally as substitutes for fluorescein. In addition, different formulations or related dyes have been developed around other fluorophores that emit light in other parts of the spectrum.

[00149] Non-radioactive and non-fluorescent label monomers are also available. For example, biotin can be attached directly to nucleotides and detected by specific and high-affinity binding to avidin or streptavidin which has been chemically coupled to an enzyme catalyzing a colorimetric reaction (such as phosphatase, luciferase, or peroxidase). Digoxigenin-labeled nucleotides can similarly be used for non-isotopic detection of nucleic acids. Biotinylated and digoxigenin-labeled nucleotides are commercially available.

[00150] Very small particles, termed nanoparticles, also can be used as label monomers to label nucleic acids. These particles range from 1-1000 nm in size and include diverse chemical structures such as gold and silver particles and quantum dots.

[00151] When irradiated with angled incident white light, silver or gold nanoparticles ranging from 40-120 nm will scatter monochromatic light with high intensity. The wavelength of the scattered light is dependent on the size of the particle. Four to five different particles in close proximity will each scatter monochromatic light which when superimposed will give a specific, unique color. The particles are being manufactured by companies such as Genicon Sciences. Derivatized silver or gold particles can be attached to a broad array of molecules including, proteins, antibodies, small molecules, receptor ligands, and nucleic acids. For example, the surface of the particle can be chemically derivatized to allow attachment to a nucleotide.

[00152] Another type of nanoparticles that can be used as a label monomer are quantum dots. Quantum dots are fluorescing crystals 1-5 nm in diameter that are excitable by a large range of wavelengths of light. These crystals emit light, such as monochromatic light, with a wavelength dependent on their chemical composition and size. Quantum dots such as CdSe, ZnSe, InP, or InAs possess unique optical properties.

[00153] Many dozens of classes of particles can be created according to the number of size classes of the quantum dot crystals. The size classes of the crystals are created either 1) by tight control of crystal formation parameters to create each desired size class of particle, or 2) by creation of batches of crystals under loosely controlled crystal formation parameters, followed by sorting according to

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1146

desired size and/or emission wavelengths. Use of quantum dots for labeling particles, in the context of the present invention, is new, but is old in the art of semiconductors. Two examples of earlier references in which quantum dots are embedded within intrinsic silicon epitaxial layers of semiconductor light emitting/detecting devices are United States Patent Nos. 5,293,050 and 5,354,707 to Chapple Sokol, *et al.*

[00154] In specific embodiments, one or more of the label attachments regions in the nanoreporter is labeled with one or more light-emitting dyes, each label attachment region containing, directly or indirectly, one or more detectable molecules or label monomers. The light emitted by the dyes can be visible light or invisible light, such as ultraviolet or infrared light. In exemplary embodiments, the dye is a fluorescence resonance energy transfer (FRET) dye; a xanthene dye, such as fluorescein and rhodamine; a dye that has an amino group in the alpha or beta position (such as a naphthylamine dye, 1-dimethylaminonaphthy1-5-sulfonate, 1-anilino-8-naphthalende sulfonate and 2-p-touidinyl-6-naphthalene sulfonate); a dye that has 3-phenyl-7-isocyanatocoumarin; an acridine, such as 9-isothiocyanatoacridine and acridine orange; a pyrene, a bensoxadiazole and a stilbene; a dye that has 3-(ε-carboxypentyl)-3'-ethy1-5,5'-dimethyloxacarbocyanine (CYA); 6-carboxy fluorescein (FAM); 5&6-carboxyrhodamine-110 (R110); 6-carboxyrhodamine-6G (R6G); N,N,N',N'-tetramethy1-6-carboxyrhodamine (TAMRA); 6-carboxy-X-rhodamine (ROX); 6-carboxy-4',5'-dichloro-2',7'-dimethoxyfluorescein (JOE); ALEXA Fluor™; Cy2; Texas Red and Rhodamine Red; 6-carboxy-2',4,7,7'-tetrachlorofluorescein (TET); 6-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (HEX); 5-carboxy-2',4',5',7'-tetrachlorofluorescein (ZOE); NAN; NED; Cy3; Cy3.5; Cy5; Cy5.5; Cy7; and Cy7.5; Alexa Fluor 350; Alexa Fluor 488; Alexa Fluor 532; Alexa Fluor 546; Alexa Fluor 568; Alexa Fluor 594; or Alexa Fluor 647.

[00155] The label monomers can be incorporated into a nanoreporter at different stages of its assembly, or into a component (*e.g.*, a complementary RNA nucleotides of the nanoreporter prior to its assembly into the nanoreporter).

[00156] A label monomer can be directly attached to a nucleotide using methods well known in the art. Nucleotides can also be chemically modified or derivatized in order to attach a label monomer. For example, a fluorescent monomer such as a fluorescein molecule can be attached to dUTP (deoxyuridine-triphosphate) using a four-atom aminoalkynyl group. Each label monomer is attached to a nucleotide making a label monomer: nucleotide complex.

[00157] This label monomer/nucleotide complex can be incorporated into nucleic acids (for example, a DNA patch or a detection oligonucleotide) in a variety of ways. For example, a label monomer/nucleotide complex can be incorporated at only one location within a nucleic acid or at two or more locations within a nucleic acid.

[00158] Amine-reactive and thiol-reactive fluorophores are available and used for labeling nucleotides and biomolecules. Generally, nucleotides are fluorescently labeled during chemical synthesis, for example, incorporation of amines or thiols during nucleotide synthesis permit addition of

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1147

fluorophores. Fluorescently labeled nucleotides are commercially available. For example, uridine and deoxyuridine triphosphates are available that are conjugated to ten different fluorophores that cover the spectrum.

[00159] A nucleotide can be attached to a label monomer first and then be incorporated into a nucleic acid. Alternatively, an existing nucleic acid can be labeled by attaching a label monomer to a nucleotide within the nucleic acid. For example aminoallyl-("AA-") modified UTP nucleotides can be incorporated into the RNA product during transcription. In various embodiments, 20% or more of UTP nucleotides in a transcription reaction to generate complementary RNA nucleotides are AA modified. In various embodiments, about 10% or about 20% to 100%, 20% to 80%, 30% to 80%, 40% to 60% or 50% to 75% of UTPs in a transcription reaction are AA-modified, in a preferred embodiment, about 40% or about 50% of UTPs in a transcription reaction are AA-modified.

[00160] In addition, for example, different types of label monomer: nucleotide complexes can be incorporated into a single acid nucleic acid, where one component of the nanoreporter code comprises more than one type of signal.

[00161] Fluorescent dyes that can be bound directly to nucleotides can also be utilized as label monomers. For example, FAM, JOE, TAMRA, and ROX are amine reactive fluorescent dyes that have been attached to nucleotides and are used in automated DNA sequencing. These fluorescently labeled nucleotides, for example, ROX-ddATP, ROX-ddCTP, ROX-ddGTP and ROX-ddUTP, are commercially available.

[00162] Other types of label monomers that may be used to label a nanoreporter are quantum dots. Due to their very small size the quantum dots can be coupled into oligonucleotides directly without affecting the solubility or use of the oligonucleotide. In a preferred embodiment, only one oligonucleotide molecule is coupled to each nanoparticle. To synthesize an oligonucleotide-nanoparticle complex in a 1:1 ratio by conventional batch chemistry, both the oligonucleotide and the nanoparticle require a single reactive group of different kinds that can be reacted with each other. For example, if an oligonucleotide has an amino group and a nanoparticle has an aldehyde group, these groups can react to form a Schiff base. An oligonucleotide can be derivatized to attach a single amino or other functional group using chemistry well known in the art. However, when a nanoparticle is derivatized, it is covered with a chemical reagent which results in coating the entire surface of the nanoparticle with several functional groups.

Affinity Tags

[00163] A variety of affinity tags known in the art may be used to purify and/or immobilize nanoreporters.

[00164] Where an affinity tag is used to immobilize a nanoreporter for the purpose of detection or imaging, it may be referred to herein as an "anchor." (See Figure 3). In some embodiments, a biotin

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1148

anchor is attached to the nanoreporter, allowing immobilization of the nanoreporter on a streptavidin coated slide.

[00165] In some embodiments, a labeled nanoreporter will contain an affinity tag at each end, A1 and A2. The labeled nanoreporter can be immobilized to a surface through the binding of A1 to an immobilized affinity partner. In the absence of an affinity binding partner for A2, the A2 end of the nanoreporter remains in solution, but in the presence of an affinity binding partner (A2'), the A2 end of the nanoreporter is also immobilized. In some embodiments, a labeled nanoreporter will contain a single affinity tag, A1. Another affinity tag, A2, can be attached to the nanoreporter by direct binding of the nanoreporter to a molecule containing A2 (*e.g.*, if the nanoreporter is or comprises a nucleic acid, it can hybridize directly with another nucleic acid to which A2 is attached). Alternatively, either affinity tag can be attached to the labeled nanoreporter via a bridging molecule, such as the bridging nucleic acid. Figure 3 illustrate yet another embodiment in which a labeled (nucleic acid-based) nanoreporter contains an affinity tag, A1, at one end. In Figure 3, the labeled nanoreporter is immobilized through the binding of A1 to an immobilized affinity partner. The other end of the nanoreporter is in solution (Figure 3A), but can be immobilized by hybridization to a complementary oligonucleotide which contains another affinity tag (A2) used to immobilize the nanoreporter (Figure 3B). Upon immobilization of A1, the nanoreporter can be stretched, or "elongated" as depicted in Figure 3, for example by electrostretching, for separation of the label attachment regions in a manner that permits detection of the nanoreporter code. Optionally, while the nanoreporter is in an elongated state, A2 is introduced and binds the end of the nanoreporter that is complementary to A2 down to the surface.

[00166] An affinity tag can be used for attachment to beads or other matrixes for a variety of useful applications including but not limited to purification.

[00167] Non-limiting examples of suitable affinity tags are provided below. It should be understood that most affinity tags could serve dual purposes: both as anchors for immobilization of the nanoreporters and tags for purification of the nanoreporters (whether fully or only partially assembled) or their components.

[00168] In certain embodiments, the affinity tag is a protein monomer. Examples of protein monomers include, but are not limited to, the immunoglobulin constant regions (*see* Petty, 1996, Metal-chelate affinity chromatography, in Current Protocols in Molecular Biology, Vol. 2, Ed. Ausubel *et al.*, Greene Publish. Assoc. & Wiley Interscience), glutathione S-transferase (GST; Smith, 1993, Methods Mol. Cell Bio. 4:220-229), the *E. coli* maltose binding protein (Guan *et al.*, 1987, Gene 67:21-30), and various cellulose binding domains (U.S. Pat. Nos. 5,496,934; 5,202,247; 5,137,819; Tomme *et al.*, 1994, Protein Eng. 7:117-123), etc.

[00169] Other affinity tags are recognized by specific binding partners and thus facilitate isolation and immobilization by affinity binding to the binding partner, which can be immobilized onto a solid support. For example, the affinity tag can be an epitope, and the binding partner an antibody.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1149

Examples of such epitopes include, but are not limited to, the FLAG epitope, the myc epitope at amino acids 408-439, the influenza virus hemagglutinin (HA) epitope, or digoxigenin ("DIG"). In other embodiments, the affinity tag is a protein or amino acid sequence that is recognized by another protein or amino acid, for example the avidin/streptavidin and biotin.

[00170] In certain aspects of the invention, the affinity tag is a nucleotide sequence. A large variety of sequences of about 8 to about 30 bases, more preferably of about 10 to about 20 bases, can be used for purification and immobilization of nanoreporters, and the sequence can be tandemly repeated (*e.g.*, from 1 to 10 tandem repeats). Such a sequence is preferably not widely represented (that is, present in fewer than 5% of the genes, more preferably, present in fewer than 3% of the genes, and, most preferably, present in fewer than 1% of the genes) in the sample being assayed (for example, where the nanoreporter is used for detection of human cellular RNA, the sequence is preferably not widely represented in the human genome); have little or no secondary structure or self-complementarity either internally or with copies of itself when multimerized (that is, all secondary structures of the multimerized tag preferably have a Tm less than 25°C at 1 M NaCI); have no significant identity or complementarity with backbone or segment sequences (that is, the Tm of complementary sequences is preferably less than 25°C at 0.2 M NaCI); and have a Tm of about 35-65°C, more preferably about 40-50°C, in 50 mM Na^+.

[00171] In certain embodiments, different sequences are used as purification and immobilization tags. In this case, for example, the purification tag can be as described above, but the immobilization tag can be in the range of 10 to 100 bases, with a Tm up to 95°C in 50 mM Na^+. An alternative embodiment would be to have the purification tag nested within the immobilization tag (*e.g.*, the affinity tag would comprise a 25-base sequence of which 15 bases are used as a purification tag and the entire 25 bases are used as the immobilization tag).

[00172] In certain instances, the affinity tag can be used for labeling a nanoreporter in addition to purifying or immobilizing the nanoreporter.

[00173] As will be appreciated by those skilled in the art, many methods can be used to obtain the coding region of the affinity tags, including but not limited to, DNA cloning, DNA amplification, and synthetic methods. Some of the affinity tags and reagents for their detection and isolation are available commercially.

Constant Region

[00174] The nanoreporters of the present invention may comprise at least one constant region. A constant region may comprise a polynucleotide sequence. In certain aspects, the polynucleotide sequence of a constant region comprises a plurality of individual repeat nucleotide sequences, such as 2, 3, 4, 5, 6, 7, 8, 9, 10, or more copies of a repeat sequence. In certain aspects, the constant region comprises four copies of a repeat sequence.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1150

[00175] In certain embodiments, an individual repeat sequence comprises about 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more nucleotide bases. In certain embodiments, an individual repeat sequences comprise about 15 nucleotide bases.

[00176] One example of a 15-base repeat sequence is 5'-GGTOTGTGTGATGTT-3' (SEQ ID NO:25). In certain embodiments, the constant region contains four copies of a 15 base repeat, such as the repeat sequence of SEQ ID NO:25.

[00177] In certain aspects, the constant region may be ligated to a nanoreporter backbone after isolation and preparation of the backbone polynucleotide sequence, as described herein. In some embodiments, the constant region represents a permanent part of the polynucleotide backbone, as if it were fused to the "coding" sequence of the backbone (*i.e.*, the constant region is cloned into the backbone). In certain embodiments, the constant region is adjacent to a restriction enzyme site for ready incorporation of the selected label attachment regions.

Target-Specific Regions

[00178] The term "target-specific sequence" refers to a molecular entity that is capable of binding a target molecule. In the context of a nanoreporter, the target-specific sequence is attached to the nanoreporter backbone.

[00179] The target specific sequence is generally an amino acid sequence (*i.e.*, a polypeptide or peptide sequence) or a nucleic acid sequence.

[00180] In specific embodiments, where the target-specific sequence is an amino acid sequence, the target-specific sequence may comprise an antibody fragment, such as an antibody Fab' fragment, a single chain Fv antibody.

[00181] The target-specific sequence is preferably a nucleic acid sequence, and is most preferably within an oligonucleotide that is either covalently attached (*e.g.*, by ligation) or noncovalently attached (*e.g.*, by hybridization) to the nanoreporter backbone. A target-specific nucleic acid sequence is preferably at least 15 nucleotides in length, and more preferably is at least 20 nucleotides in length. In specific embodiments, the target-specific sequence is about 10 to 500, 20 to 400, 30 to 300, 40 to 200, or 50 to 100 nucleotides in length. In other embodiments, the target-specific sequence is about 30 to 70, 40 to 80, 50 to 90, or 60 to 100, 30 to 120, 40 to 140, or 50 to 150 nucleotides in length.

[00182] A target-specific nucleotide sequence preferably has a Tm of about 65-90°C for each probe in 825 mM Na^+ (5X SSC), most preferably about 78-83°C.

[00183] In certain preferred embodiments, the target specific sequence of each probe of a dual nanoreporter is about 35 to 100 nucleotides (for a total target sequence of about 70 to 200 nucleotides, covered by 2 probes), most preferably about 40 to 50 nucleotides for each probe (for a total of about 80 to 100 nucleotides).

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1151

Target Molecules

[00184] The term "target molecule" is the molecule detected or measured by binding of a labeled nanoreporter whose target-specific sequence(s) recognize (*i.e.*, are specific binding partners) thereto. Preferably, a target molecule can be, but is not limited to, any of the following: nucleic acid, peptide, a polypeptide/protein (*e.g.*, a bacterial or viral protein or an antibody), a lipid, a carbohydrate, a glycoprotein, a glycolipid, a small molecule, an organic monomer, or a drug. Nucleic acids that can be analyzed by the methods herein include: double-stranded DNA, single-stranded DNA, single-stranded DNA hairpins, DNA/RNA hybrids, RNA (e.g. mRNA or miRNA) and RNA hairpins. Generally, a target molecule is a naturally occurring molecule or a cDNA of a naturally occurring molecule or the complement of said cDNA.

[00185] A target molecule can be part of a biomolecular sample that contains other components or can be the sole or major component of the sample. A target molecule can be a component of a whole cell or tissue, a cell or tissue extract, a fractionated lysate thereof or a substantially purified molecule. The target molecule can be attached in solution or solid-phase, including, for example, to a solid surface such as a chip, microarray or bead. Also the target molecule can have either a known or unknown structure or sequence.

[00186] In certain specific embodiments, that target molecule is not a chromosome. In other specific embodiments, the target molecule is no greater than 1,000 kb (or 1 mb) in size, no greater than 500 kb in size, no greater than 250 kb in size, no greater than 175 kb in size, no greater than 100 kb in size, no greater than 50 kb in size, no greater than 20 kb in size, or no greater than 10 kb in size. In yet other specific embodiments, the target molecule is isolated from its cellular milieu.

Biomolecular Samples

[00187] The nanoreporter systems of the invention can be used to detect target molecule in any biomolecular sample. As will be appreciated by those in the art, the sample may comprise any number of things, including, but not limited to: biological samples, such as cells (including both primary cells and cultured cell lines), cell lysates, or extracts (including but not limited to RNA extracts; purified mRNA), tissues and tissue extracts (including but not limited to RNA extracts; purified mRNA); bodily fluids (including, but not limited to, blood, urine, serum, lymph, bile, cerebrospinal fluid, interstitial fluid, aqueous or vitreous humor, colostrum, sputum, amniotic fluid, saliva, anal and vaginal secretions, perspiration and semen, a transudate, an exudate (*e.g.*, fluid obtained from an abscess or any other site of infection or inflammation) or fluid obtained from a joint (*e.g.*, a normal joint or a joint affected by disease such as rheumatoid arthritis, osteoarthritis, gout or septic arthritis) of virtually any organism, with mammalian samples being preferred and human samples being particularly preferred; environmental samples (including, but not limited to, air, agricultural, water and soil samples); biological warfare agent samples; research samples including

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1152

extracellular fluids, extracellular supernatants from cell cultures, inclusion bodies in bacteria, cellular compartments, cellular periplasm, mitochondria compartment, etc.

[00188] The biomolecular samples can be indirectly derived from biological specimens. For example, where the target molecule of interest is a cellular transcript, *e.g.*, a messenger RNA, the biomolecular sample of the invention can be a sample containing cDNA produced by a reverse transcription of messenger RNA. In another example, the biomolecular sample of the invention is generated by subjecting a biological specimen to fractionation, *e.g.*, size fractionation or membrane fractionation.

[00189] The biomolecular samples of the invention may be either native, *e.g.*, not subject to manipulation or treatment, or treated, which can include any number of treatments, including exposure to candidate agents including drugs, genetic engineering (*e.g.*, the addition or deletion of a gene), etc.

[00190] Biomolecular samples may also include environmental samples, such as those containing bacteria or other organisms, such as diatoms, dinoflagellates, algae, among others, such as in certain marine or earth-based samples.

Methods

[00191] In certain aspects the invention provides methods for detection and/or quantification of one or more target molecules. In particular, the invention provides stable nanoreporters that are capable of binding individual target molecules, and provide improved detection of target molecules. Through nanoreporters' label codes, the binding of the nanoreporter probes to target molecules results in the identification of the target molecules. The detection and/or quantification can be performed using a single nanoreporter probe system or a dual nanoreporter probe system as described herein.

[00192] In some embodiments, the invention provides methods of detecting the presence of a specific target molecule in a biomolecular sample comprising contacting the sample with a nanoreporter probe as described herein under conditions that allow binding of the target-specific sequences in the probes to a target molecule in the sample, and detecting the signal associated with the uniquely labeled nanoreporter probe. In some embodiments, the biomolecular sample is a biological sample. In some embodiments, the biological sample is selected from a cell, a cell lysate, a tissue sample, a tissue extract, or a bodily fluid. In some embodiments, the biomolecular sample is an environmental sample.

[00193] In some embodiments, the invention provides methods of detecting the presence of a plurality of specific target molecules in a biomolecular sample comprising contacting the sample with a plurality of nanoreporter probes as described herein under conditions that allow binding of the target-specific sequences in the probes to a target molecule in the sample, and detecting the signal associated with the uniquely labeled nanoreporter probes.

[00194] In some embodiments, the invention provides methods for determining the presence of at least one target molecule in a sample, comprising forming at least one molecular complex comprising (a) at

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1153

least one target molecule and (b) at least one nanoreporter probe comprising an unique target-specific region and a region comprising an unique, designed nanoreporter wherein said nanoreporter comprises a plurality of different detectable molecules, and individually detecting said at least one molecular complex or at least part of said at least one molecular complex to determine the presence of at least one target molecule in the sample by a method comprising individually counting the presence of one or more nanoreporter molecules. In some embodiments the methods of the invention provides for determining the presence of a plurality of target molecules by a method comprising forming a plurality molecular complexes, each complex comprising (a) at least one target molecule and (b) at least one nanoreporter probe comprising an unique target-specific region and a region comprising an unique, designed nanoreporter, wherein each nanoreporter comprises a plurality of different detectable molecules; and wherein each nanoreporter probe comprises a different nanoreporter region. In some embodiments, the number of molecular counts above background of said molecular complex after normalization of the sample is higher than 300, 400, 450, 500, 600, 700, 800, 900 or 1000 molecular counts. In some embodiments, the number of molecular counts above background of the molecular complex after normalization of the sample is higher than 400 molecular counts. In some embodiments, the percentage of valid molecular counts of the nanoreporter probes is higher that about 10, 11, 12, 12.5, 14, 15, 20, 25, 30, 35, 40, 50, 60, 70, 80 or 90 %. In some embodiments, the percentage of valid molecular counts of the nanoreporter probes is higher that about 12.5 %. In some embodiments, the number of molecular counts above background of the nanoreporter described herein after normalization of said sample is at least 6 fold higher than a comparable nanoreporter comprising M13 DNA. In some embodiments, the number of molecular counts above background of the nanoreporter described herein after normalization of said sample is at least 2 fold higher than a comparable nanoreporter comprising M13 DNA.

[00195] In some embodiments, the invention provides methods for determining the presence of at least one target molecule in a sample, comprising (1) forming at least one molecular complex comprising (a) at least one target molecule and (b) at least one probe comprising an unique target-specific region and a region comprising an unique, designed nanoreporter, wherein said nanoreporter comprises a single-stranded nucleic acid backbone, said backbone comprising a plurality of label attachment regions covalently attached together in an unique linear combination, wherein each label attachment region is hybridized to an unique complementary polynucleotide sequence having attached thereto one or more detectable molecules, and (2) individually detecting the at least one molecular complex or at least part of said at least one molecular complex to determine the presence of at least one target molecule in the sample. In some embodiments, the percentage of valid molecular counts of the nanoreporter probes is higher that about 10, 11, 12, 12.5, 14, 15, 20, 25, 30, 35, 40, 50, 60, 70, 80 or 90 %. In some embodiments, the percentage of valid molecular counts of the nanoreporter probes is higher that about 12.5 %.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1154

[00196] In some embodiments, the number of molecular counts above background of the nanoreporter described herein after normalization of said sample is at least 2 fold higher than a nanoreporter comprising M13 DNA. In some embodiments, the Tm of the complementary polynucleotide sequences when hybridized to its label attachment regions are higher than the Tm of a polynucleotide sequences complementary to a M13 DNA template when hybridized to a nanoreporter probe comprising the M13 DNA; The M13 nanoreporter probe comprises the same target specific sequence of the nanoreporter probe, and comprises a single–stranded backbone comprising a plurality of M13 DNA regions covalently attached together wherein each region is hybridized to a complementary polynucleotide sequence having attached thereto one or more detectable molecules. In some embodiments, the invention provide methods of determining the presence of a plurality of target molecules by a method comprising forming a plurality molecular complexes, each complex comprising (a) at least one target molecule and (b) at least one nanoreporter probe, where each nanoreporter probe has a higher melting temperature than a comparable nanoreporter comprising M13 DNA.

[00197] Examples of specific embodiments of the invention are shown in Figures 1 and 2.

[00198] In addition to detecting an overall signal generated from a labeled nanoreporter, the invention provides for the determination of the spatial location of signals emanating from the label monomers (*e.g.*, spots) on a nanoreporter, each spot representing the aggregate signal from label monomers attached to a given label attachment region. A spot may contain signals of the same wavelength or of different wavelengths. Thus, the nature of the spots on a nanoreporter and their location constitutes the nanoreporter code.

[00199] In some embodiments, the label nanoreporter is captured and stretched in a location (See Figure 4). In some embodiments, prior to stretching a nanoreporter, the nanoreporter is immobilized to a solid surface using an affinity tag, as described above (See Figure 3). In certain aspects of the invention, one end of a nanoreporter is immobilized, either through specific or non-specific binding to a solid surface, the nanoreporter is stretched, and then the other end of the reporter is immobilized, also either through specific or non-specific binding to a solid surface (See Figure 3). Methods for capturing, stretching and immobilization of nucleic acids are well known in the art. Examples of methods that can be used in the methods described herein are described in US patent 7,473,767 entitled "Methods for detection and quantification of analytes in complex mixtures", US patent publication no. 2007/0166708 entitled "Methods for detection and quantification of analytes in complex mixtures", US application number 11/645,270 entitled "Compositions comprising oriented, immobilized macromolecules and methods for their preparation", PCT application no US06/049274 entitled "Nanoreporters and methods of manufacturing and use thereof", and US provisional application 60/088,988 entitled "Stable nanoreporter", all of which are incorporated by reference herein in its entirety.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1155

[00200] Thus, in some embodiments, the nanoreporter is "frozen" in its stretched, or extended, state, to facilitate resolution of the nanoreporters code by detecting and/or imaging the signals provided (*e.g.*, emitted) by the label monomers attached to a nanoreporter and their locations relative to one another. These aspects of the invention are described in US patent 7,473,767 entitled "Methods for detection and quantification of analytes in complex mixtures", US patent publication no. 2007/0166708 entitled "Methods for detection and quantification of analytes in complex mixtures", US application number 11/645,270 entitled "Compositions comprising oriented, immobilized macromolecules and methods for their preparation", PCT application no US06/049274 entitled "Nanoreporters and methods of manufacturing and use thereof", and US provisional application 60/088,988 entitled "Stable nanoreporter", all of which are incorporated by reference herein in its entirety.

[00201] In some embodiments, the nanoreporter can be immobilized to the substrate. In the methods of the invention, the substrate for immobilization can be any substrate capable of selectively binding the nanoreporter apparent to those of skill in the art. In some embodiments, a first portion of the nanoreporter can be immobilized to the substrate via an avidin-biotin binding pair. In certain embodiments, the nanoreporter can comprise a biotin moiety in its first portion. For instance, a polynucleotide nanoreporter can comprise a biotinylated nucleotide residue. In a dual nanoreporter probe system, one of the nanoreporter probes (e.g. capture nanoreporter probe) can comprise a biotin moiety. This probe might be labeled or unlabeled. In preferred embodiments, when a dual nanoreporter system is used, the nanoreporter probe comprising a biotin moiety is unlabeled. The substrate comprising avidin can be any substrate comprising avidin known to those of skill in the art. Useful substrates comprising avidin are commercially available including TB0200 (Accelr8), SAD6, SAD20, SAD100, SAD500, SAD2000 (Xantec), SuperAvidin (Array-It), streptavidin slide (catalog #MPC 000, Xenopore) and STREPTAVIDINnslide (catalog #439003, Greiner Bio-one).

[00202] In certain embodiments, the first portion of the nanoreporter can comprise a nucleotide sequence that is capable of selectively binding a nucleotide sequence on the substrate (e.g. biotin). In certain aspects, the constant region (*e.g.*, multiple copies of a 15-base repeat sequence) of the nanoreporter is capable of selectively binding a complementary polynucleotide sequence that is immobilized on the substrate. In a specific embodiment, if the nanoreporter comprises four copies of the exemplary 15-base repeat sequence of 5'-GGTCTGTGTGATGTT-3' (SEQ ID NO:25), an oligonucleotide having the sequence 5'-AACATCACACAGACC AACATCACACAGACC AACATCACACAGACC AACATCACACAGACC AGCCCTTTG-3' (SEQ ID NO:26) may be utilized to immobilize the nanoreporter to a substrate.

[00203] In some embodiments where a dual nanoreporter probe system is used, a first unlabeled capture nanoreporter probe comprising biotin and a second labeled nanoreporter probe comprising a constant region are used for the detection and/or quantification of a target molecule (See Figures 1 and 2). Both nanoreporter probes bind to the target molecule via their target specific sequence

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1156

forming a complex comprising the target molecule and the nanoreporter probes. The complex is then captured and immobilized into an avidin containing substrate such as the ones described herein. The complex can then be stretched and extended as described herein. The constant region in the nanoreporter probe can then be bound to a complementary polynucleotide sequence that is immobilized on the substrate. Examples of methods that can be used in the methods described herein with respect to dual nanoreporter systems are described in PCT application no US06/049274 entitled "Nanoreporters and methods of manufacturing and use thereof", and US provisional application 60/088,988 entitled "Stable nanoreporter", all of which are incorporated by reference herein in its entirety.

[00204] Examples of specific embodiments of the invention are shown in Figures 2 – 4.

[00205] In certain aspects, the invention provides methods for preparing a suitable population of unique nucleic acid backbones, comprising (a) selecting a set of label attachment regions from a library of unique, designed label attachment regions, and (b)covalently attaching the set of label attachment regions to each other in a linear combination that is different from the other backbones in the population, and repeating steps (a)-(b) until a suitable population of unique nucleic acid backbones has been prepared. In some embodiments the label attachment regions comprise about 800 to 1300 nucleotide bases. In some embodiments the label attachment regions have a G/C content of about 50%. In some embodiments, adenine bases are spaced on average at about every 8 to 16 bases in the label attachment regions. In some embodiments, the label attachment regions are devoid of secondary structures. In some embodiments, each selected label attachment region is different from the other selected label attachment regions in the set. In some embodiments, the selecting step (a) comprises selecting a set of 3, 4, 5, 6, 7, 8 label attachment regions. In some embodiments, the selecting step (a) comprises selecting a set of 6 label attachment regions. In some embodiments, the label attachment regions are hybridized to a complementary RNA polynucleotide having attached thereto a detectable molecule. In some embodiments, the complementary RNA polynucleotide has a G/C ratio of at least 1/1. In some embodiments, the complementary RNA polynucleotide has a G/C ratio of about 3/2.

[00206] In certain aspects, the invention provides methods for preparing a suitable population of unique nucleic acid backbones, comprising (a) selecting a set of label attachment regions from a library of unique, designed label attachment regions, wherein each label attachment region comprises about 800 to 1300 nucleotide bases and has a G/C content of about 50%, wherein each label attachment region comprises a regularly-spaced pattern of adenine bases, and (b)covalently attaching the set of label attachment regions to each other in a linear combination that is different from the other backbones in the population, and repeating steps (a)-(b) until a suitable population of unique nucleic acid backbones has been prepared. In some embodiments, adenine bases are spaced at about every 8 to 16 bases in the label attachment regions. In some embodiments, the label attachment regions are devoid of secondary structures. In some embodiments, each selected label attachment region is different from the other selected label attachment regions in the set. In some embodiments, the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1157

selecting step (a) comprises selecting a set of 3, 4, 5, 6, 7, 8 label attachment regions. In some embodiments, the selecting step (a) comprises selecting a set of 6 label attachment regions. In some embodiments, the label attachment regions are hybridized to a complementary RNA polynucleotide having attached thereto a detectable molecule. In some embodiments, the complementary RNA polynucleotide has a G/C ratio of about 3/2.

Methods of Using Selectively Immobilized, Extended, or Oriented Nanoreporters

[00207] In certain embodiments, the invention provides selectively immobilized, elongated nanoreporters that can be used to create macromolecular barcodes for the purposes of separation and sequential detection of labels. These labels spaced along the molecule provide a unique code that can be read when the nanoreporter is extended and immobilized. Extension and selective immobilization can facilitate the decoding of the macromolecular barcode.

[00208] The selectively immobilized, elongated nanoreporters can be used in any context where detection or imaging of a nanoreporter might be useful. They can be used for diagnostic, prognostic therapeutic and screening purposes. For instance, they can be applied to the analysis of biomolecular samples obtained or derived from a patient so as to determine whether a diseased cell type is present in the sample and/or to stage the disease. They can be used to diagnose pathogen infections, for example infections by intracellular bacteria and viruses, by determining the presence and/or quantity of markers of bacterium or virus, respectively, in the sample. The compositions and methods of the invention can be used to quantitate target molecules whose abundance is indicative of a biological state or disease condition, for example, blood markers that are upregulated or downregulated as a result of a disease state. In addition, the compositions and methods of the invention can be used to provide prognostic information that assists in determining a course of treatment for a patient.

Detection of Nanoreporters

[00209] Nanoreporters are detected by any means available in the art that is capable of detecting the specific signals on a given nanoreporter. Where the nanoreporter is fluorescently labeled, suitable consideration of appropriate excitation sources may be investigated. Possible sources may include but are not limited to arc lamp, xenon lamp, lasers, light emitting diodes or some combination thereof. The appropriate excitation source is used in conjunction with an appropriate optical detection system, for example an inverted fluorescent microscope, an epi-fluorescent microscope or a confocal microscope. Preferably, a microscope is used that can allow for detection with enough spatial resolution to determine the sequence of the spots on the nanoreporter. For example in one embodiment an image of a dual nanoreporter hybridized to a target molecule can be obtained. If for example, the nanoreporters are labeled with three different colors, Alexa 488, Cy3 and Alexa 647 (labeled 1, 2 and 3, respectively). Colors 1, 2 and 3 are each acquired in different channels and the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1158

first and second registers, which can be seen as rows of spots, are shifted up by several pixels to be able to show each register individually.

[00210] Examples of methods for detection of nanoreporters that can be used in the methods of the invention are described in US patent 7,473,767 entitled "Methods for detection and quantification of analytes in complex mixtures", US patent publication no. 2007/0166708 entitled "Methods for detection and quantification of analytes in complex mixtures", US application number 11/645,270 entitled "Compositions comprising oriented, immobilized macromolecules and methods for their preparation", PCT application no US06/049274 entitled "Nanoreporters and methods of manufacturing and use thereof", and US provisional application 60/088,988 entitled "Stable nanoreporter", all of which are incorporated by reference herein in its entirety.

Microscope and Objective Lens Selection

[00211] The major consideration regarding the microscope objective lens is with the optical resolution, which is determined by its numerical aperture (NA). Generally, the larger the NA, the better the optical resolution. The required NA is preferably at least 1.07 based on the relationship of $\delta = 0.61\lambda/NA$ (δ = optical resolution and λ = wavelength). The amount of light that is collected by an objective is determined by NA^4/Mag^2 (Mag = magnification of the objective). Therefore, in order to collect as much light as possible, objectives with high NA and low magnifications should be selected.

CCD Camera Selection and Image Capture Techniques

[00212] When selecting a CCD camera, the first consideration is the pixel size, which partially determines the final resolution of the imaging system. Optimally the optical resolution should not be compromised by the CCD camera. For example, if the optical resolution is 210-300 nm, which corresponds to 12.6-18 μm on a CCD chip after a 60 x magnification, in order to resolve and maintain the optical resolution there should be at least two pixels to sample each spot. Or the pixel size of the CCD chip should be at most 6.3-9 μm.

[00213] The second consideration is detection sensitivity which can be determined by many factors that include but are not limited to pixel size, quantum efficiency, readout noise and dark noise. To achieve high sensitivity, select a qualitative camera with big pixel size (which can give big collection area), high quantum efficiency and low noise. An exemplary camera with these criteria is the Orca-Ag camera from Hamamatsu Inc. The chip size is 1344 x 1024 pixels; when using the 60x objective, the field of view is 144 x 110 μm^2.

Applications of Nanoreporter Technology

[00214] The compositions and methods of the invention can be used for diagnostic, prognostic therapeutic and screening purposes. The present invention provides the advantage that many different target molecules can be analyzed at one time from a single biomolecular sample using the methods of the invention. This allows, for example, for several diagnostic tests to be performed on one sample.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1159

[00215] The methods described herein discriminate between nucleotide sequences. The difference between the target nucleotide sequences can be, for example, a single nucleic acid base difference, a nucleic acid deletion, a nucleic acid insertion, or rearrangement. Such sequence differences involving more than one base can also be detected. In some embodiments, the oligonucleotide probe sets have substantially the same length so that they hybridize to target nucleotide sequences at substantially similar hybridization conditions. As a result, the process of the present invention is able to detect infectious diseases, genetic diseases, and cancer. It is also useful in environmental monitoring, forensics, and food science. Examples of genetic analyses that can be performed on nucleic acids include e-g., SNP detection, STR detection, RNA expression analysis, promoter methylation, gene expression, virus detection, viral subtyping and drug resistance.

[00216] In the area of environmental monitoring, the present invention can be used for detection, identification, and monitoring of pathogenic and indigenous microorganisms in natural and engineered ecosystems and microcosms such as in municipal waste water purification systems and water reservoirs or in polluted areas undergoing bioremediation. It is also possible to detect plasmids containing genes that can metabolize xenobiotics, to monitor specific target microorganisms in population dynamic studies, or either to detect, identify, or monitor genetically modified microorganisms in the environment and in industrial plants.

[00217] The present invention can also be used in a variety of forensic areas, including for human identification for military personnel and criminal investigation, paternity testing and family relation analysis, HLA compatibility typing, and screening blood, sperm, or transplantation organs for contamination.

[00218] In the food and feed industry, the present invention has a wide variety of applications. For example, it can be used for identification and characterization of production organisms such as yeast for production of beer, wine, cheese, yogurt, bread, etc. Another area of use is with regard to quality control and certification of products and processes (e.g., livestock, pasteurization, and meat processing) for contaminants. Other uses include the characterization of plants, bulbs, and seeds for breeding purposes, identification of the presence of plant-specific pathogens, and detection and identification of veterinary infections.

Diagnostic/Prognostic Methods

[00219] The present methods can be applied to the analysis of biomolecular samples obtained or derived from a patient so as to determine whether a diseased cell type is present in the sample and/or to stage the disease.

[00220] In some embodiments, the methods described herein are used in the diagnosis of a condition. As used herein the term “diagnose” or “diagnosis” of a condition includes predicting or diagnosing the condition, determining predisposition to the condition, monitoring treatment of the condition, diagnosing a therapeutic response of the disease, and prognosis of the condition, condition progression, and response to particular treatment of the condition. For example, a blood sample can

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1160

be assayed according to any of the methods described herein to determine the presence and/or quantity of markers of a cancerous cell type in the sample, thereby diagnosing or staging the cancer.

[00221] Cancers which can be detected by the process of the present invention generally involve oncogenes, tumor suppressor genes, or genes involved in DNA amplification, replication, recombination, or repair. Examples of these include: BRCA1 gene, p53 gene, APC gene, Her2/Neu amplification, Bcr/Ab1, K-ras gene, and human papillomavirus Types 16 and 18. Various aspects of the present invention can be used to identify amplifications, large deletions as well as point mutations and small deletions/insertions of the above genes in the following common human cancers: leukemia, colon cancer, breast cancer, lung cancer, prostate cancer, brain tumors, central nervous system tumors, bladder tumors, melanomas, liver cancer, osteosarcoma and other bone cancers, testicular and ovarian carcinomas, head and neck tumors, and cervical neoplasms.

[00222] Genetic diseases can also be detected by the process of the present invention. This can be carried out by prenatal or post-natal screening for chromosomal and genetic aberrations or for genetic diseases. Examples of detectable genetic diseases include: 21 hydroxylase deficiency, cystic fibrosis, Fragile X Syndrome, Turner Syndrome, Duchenne Muscular Dystrophy, Down Syndrome or other trisomies, heart disease, single gene diseases, HLA typing, phenylketonuria, sickle cell anemia, Tay-Sachs Disease, thalassemia, Klinefelter Syndrome, Huntington Disease, autoimmune diseases, lipidosis, obesity defects, hemophilia, inborn errors of metabolism, and diabetes.

[00223] Alternatively, the methods described herein can be used to diagnose pathogen infections, for example infections by intracellular bacteria and viruses, by determining the presence and/or quantity of markers of bacterium or virus, respectively, in the sample.

[00224] A wide variety of infectious diseases can be detected by the process of the present invention. Typically, these are caused by bacterial, viral, parasite, and fungal infectious agents. The resistance of various infectious agents to drugs can also be determined using the present invention.

[00225] Bacterial infectious agents which can be detected by the present invention include Escherichia coli, Salmonella, Shigella, Klebsiella, Pseudomonas, Listeria monocytogenes, Mycobacterium tuberculosis, Mycobacterium aviumintracellulare, Yersinia, Francisella, Pasteurella, Brucella, Clostridia, Bordetella pertussis, Bacteroides, Staphylococcus aureus, Streptococcus pneumonia, B-Hemolytic strep., Corynebacteria, Legionella, Mycoplasma, Ureaplasma, Chlamydia, Neisseria gonorrhea, Neisseria meningitides, Hemophilus influenza, Enterococcus faecalis, Proteus vulgaris, Proteus mirabilis, Helicobacter pylori, Treponema palladium, Borrelia burgdorferi, Borrelia recurrentis, Rickettsial pathogens, Nocardia, and Acitnomycetes.

[00226] Fungal infectious agents which can be detected by the present invention include Cryptococcus neoformans, Blastomyces dermatitidis, Histoplasma capsulatum, Coccidioides immitis, Paracoccidioides brasiliensis, Candida albicans, Aspergillus fumigautus, Phycomycetes (Rhizopus), Sporothrix schenckii, Chromomycosis, and Maduromycosis.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1161

[00227] Viral infectious agents which can be detected by the present invention include human immunodeficiency virus, human T-cell lymphocytotrophic virus, hepatitis viruses (e.g., Hepatitis B Virus and Hepatitis C Virus), Epstein-Barr Virus, cytomegalovirus, human papillomaviruses, orthomyxo viruses, paramyxo viruses, adenoviruses, corona viruses, rhabdo viruses, polio viruses, toga viruses, bunya viruses, arena viruses, rubella viruses, and reo viruses.

[00228] Parasitic agents which can be detected by the present invention include Plasmodium falciparum, Plasmodium malaria, Plasmodium vivax, Plasmodium ovale, Onchoverva volvulus, Leishmania, Trypanosoma spp., Schistosoma spp., Entamoeba histolytica, Cryptosporidum, Giardia spp., Trichimonas spp., Balatidium coli, Wuchereria bancrofti, Toxoplasma spp., Enterobius vermicularis, Ascaris lumbricoides, Trichuris trichiura, Dracunculus medinesis, trematodes, Diphyllobothrium latum, Taenia spp., Pneumocystis carinii, and Necator americanis.

[00229] The present invention is also useful for detection of drug resistance by infectious agents. For example, vancomycin-resistant Enterococcus faecium, methicillin-resistant Staphylococcus aureus, penicillin-resistant Streptococcus pneumoniae, multi-drug resistant Mycobacterium tuberculosis, and AZT-resistant human immunodeficiency virus can all be identified with the present invention

[00230] Thus, the target molecules detected using the compositions and methods of the invention can be either patient markers (such as a cancer marker) or markers of infection with a foreign agent, such as bacterial or viral markers.

[00231] Because of the quantitative nature of nanoreporters, the compositions and methods of the invention can be used to quantitate target molecules whose abundance is indicative of a biological state or disease condition, for example, blood markers that are upregulated or downregulated as a result of a disease state.

[00232] In addition, the compositions and methods of the invention can be used to provide prognostic information that assists in determining a course of treatment for a patient. For example, the amount of a particular marker for a tumor can be accurately quantified from even a small sample from a patient. For certain diseases like breast cancer, overexpression of certain genes, such as Her2-neu, indicate a more aggressive course of treatment will be needed.

[00233]

Analysis of Pathology Samples

[00234] RNA extracted from formaldehyde-or paraformaldehyde-fixed paraffin-embedded tissue samples is typically poor in quality (fragmented) and low in yield. This makes gene expression analysis of low-expressing genes in histology samples or archival pathology tissues extremely difficult and often completely infeasible. The nanoreporter technology can fill this unmet need by allowing the analysis of very small quantities of low-quality total RNA.

[00235] To use nanoreporter technology in such an application, total RNA can be extracted from formaldehyde-or paraformaldehyde-fixed paraffin-embedded tissue samples (or similar) using commercially available kits such as RecoverAll Total Nucleic Acid Isolation Kit (Ambion) following

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1162

manufacturer's protocols. RNA in such samples is frequently degraded to small fragments (200 to 500 nucleotides in length), and many paraffin-embedded histology samples only yield tens of nanograms of total RNA. Small amounts (5 to 100 ng) of this fragmented total RNA can be used directly as target material in a nanoreporter hybridization following the assay conditions described herein.

Screening Methods

[00236] The methods of the present invention can be used, inter alia, for determining the effect of a perturbation, including chemical compounds, mutations, temperature changes, growth hormones, growth factors, disease, or a change in culture conditions, on various target molecules, thereby identifying target molecules whose presence, absence or levels are indicative of particular biological states. In a preferred embodiment, the present invention is used to elucidate and discover components and pathways of disease states. For example, the comparison of quantities of target molecules present in a disease tissue with "normal" tissue allows the elucidation of important target molecules involved in the disease, thereby identifying targets for the discovery/screening of new drug candidates that can be used to treat disease.

Kits Comprising Nanoreporters

[00237] The invention further provides kits comprising one or more components of the invention. The kits can comprise, for example, a substrate according to the invention and one or more extended or oriented, or both, nanoreporters selectively immobilized on the substrate. The kits can be used for any purpose apparent to those of skill in the art, including those described above.

[00238] In certain embodiments, the invention provides kits for preparing at least one uniquely labeled nanoreporter comprising at least three label attachment regions each comprising about 800 to 1300 nucleotide bases, a G/C content of about 50%; and at least three complementary polynucleotide sequences having attached thereto a detectable molecule, wherein the complementary polynucleotide sequences have a G/C ratio of at least 1/1. In some embodiments, the complementary polynucleotide sequences have a G/C ratio of about 3/2. The kit can further comprise at least three target specific probes.

[00239] In certain embodiments, the present invention also provides kits useful for the extension and selective immobilization of nanoreporters. The kits can comprise a substrate for immobilization and one or more binding partners to facilitate extension or immobilization of a nanoreporter. The binding partners could, in certain embodiments, comprise a moiety useful for extension of the nanoreporter in an appropriate force. In certain embodiments, the binding partners could facilitate immobilization or selective immobilization of the nanoreporter to the surface. In further embodiments, the kit could comprise a nanoreporter for extension and immobilization. In further embodiments, the kit could comprise a device capable of extending the nanoreporter.

[00240] The kits can contain a population of nanoreporters as described herein.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1163

[00241] The kits can contain pre-labeled nanoreporters, or unlabeled nanoreporters with one or more components for labeling the nanoreporters. Moreover, the nanoreporters provided in a kit may or may not have target-specific sequences pre-attached. In one embodiment, the target sequences are provided in the kit unattached to the nanoreporter backbone.

[00242] The kit can include other reagents as well, for example, buffers for performing hybridization reactions, linkers, restriction endonucleases, and DNA I ligases.

[00243] The kit also will include instructions for using the components of the kit, and/or for making and/or using the labeled nanoreporters.

[00244] The present invention may be further understood by the non-limiting examples provided below.

EXAMPLES

Example 1: Design and Manufacturing Protocol for a De Novo 1 (DV1) Nanoreporter Backbone Library

[00245] To construct a library of diverse Nanoreporter backbones, label attachment regions were selected from a library of unique, rationally designed polynucleotide sequences and cloned in various combinations into a plasmid vector having the polynucleotide sequence set forth in SEQ ID NO:27. This vector sequence does not end up in the final reporter backbone, but is utilized for cloning and propagating the backbone sequences.

[00246] Specifically, each unique, code-specific DV1 plasmid was constructed of six approximately 1100 base pair label attachment regions, which were selected from the polynucleotide sequences of SEQ ID NOS:1-24. The label attachment regions defined by these polynucleotide sequences were cloned in various combinations into the vector of SEQ ID NO:27, each label attachment region corresponding to a fixed position on a given backbone, *i.e.*, position 1, position 2, through position 6. The label attachment regions at each position in a given backbone were different from the other label attachment regions in that same backbone. The 3' end of each position-6 sequence was cloned next to 4 copies of a common 15-base repeat known as the G-4 repeat (SEQ ID NO:26), used in purification and immobilization of the reporters. 972 nanoreporter backbones, each having a unique linear combination of label attachment regions, were designed and cloned according to routine molecular biological techniques.

[00247] When each label attachment region is assigned one of four detectable molecules, the use of six-label attachment regions in a four-color reporter system provides a possible 4096 unique nanoreporters. In this Example, SEQ ID NOS:1, 5, 9, 13, 17, and 21 were assigned a blue fluorophore, SEQ ID NOS:2, 6, 10, 14, 18, and 22 were assigned a green fluorophore, SEQ ID NOS:3, 7, 11, 15, 19, and 23 were assigned a yellow fluorophore, and SEQ ID NOS: 4, 8, 12, 16, 20, and 24 were assigned a red fluorophore. Thus, in this Example, even if a given backbone comprises 6 label attachment regions assigned a blue fluorophore (*e.g.*, SEQ ID NOS:1, 5, 9, 13, 17, and 21 in positions

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1164

1-6), each individual label attachment region would have a different polynucleotide sequence than the other label attachment regions in that same backbone.

[00248] The individual plasmids were amplified in bacteria, isolated, and converted into linear, single-stranded backbones. Specifically, linear single-stranded DNA backbones were made from a double stranded plasmid DNA using a four step protocol: (i) dsDNA was linearized with a restriction enzyme, (ii) linearized DNA was dephosphorylated with a thermolabile phosphatase, (iii) the DNA was digested with a second restriction enzyme to separate the cloning vector from the backbone sequence, and (iv) the mixture was digested with a strand-specific lambda exonuclease digestion, leaving only one strand of the backbone fragment intact.

[00249] The single stranded backbones were ligated to target-specific nucleotide sequences and incubated with dye-colored complementary RNA nucleotide sequences to make labeled nanoreporters.

Example 2: Generating Dye-Colored Complementary RNA Polynucleotides

[00250] *In vitro* transcription (IVT) reactions were utilized to generate amino-allyl modified complementary RNA polynucleotides using MEGAscript T3, T7 or SP6 kits (Ambion®), following the manufacturer's instructions with the following specifics and modifications.

[00251] Plasmids containing an RNA polymerase promoter and the polynucleotide sequence of interest were linearized by restriction digestion, ethanol-precipitated and used as templates. In this example, the sequences set forth in SEQ ID NOS:1-24 were used as templates in the IVT reaction, generating 24 unique complementary RNA polynucleotides.

[00252] Amino-allyl-UTP (aaUTP) (Fermentas) was substituted for the UTP supplied in the kit, or a mixture of aaUTP and the supplied UTP was used. The level of dye-incorporation in a complementary RNA polynucleotide correlates with the number of possible amino-allyl (aa) attachment sites. To make the brightest possible colored segments, 100% aaUTP should be used. To change the brightness, a mixture of aaUTP and unmodified UTP in any ratio can be used in the IVT reactions to modify the number of aa sites present on a complementary RNA polynucleotide. For example, to incorporate dyes at 50% of the regularly repeated base, a 1:1 mixture of aaUTP and UTP was utilized. The final concentration of aaUTP+UTP together was the same as that of each of the other three nucleotides (*i.e.*, ATP, CTP, GTP).

[00253] The IVT reaction was allowed to proceed at 37°C for 22 hours. Following the reaction, the amino-allyl modified RNA transcripts were purified using an RNeasy kit (Qiagen®) following the manufacturer's instructions.

[00254] The dye coupling reaction described below is for a 1 mg reaction. Ethanol-precipitation was performed on 1 mg of an amino-allyl (aa) modified RNA polynucleotide. Solid dyes (Alexa 488 TFP ester, Alexa 546 succinimidyl ester, Alexa 594 succinimidyl ester, and Alexa 647 succinimidyl ester; Invitrogen) were resuspended at 11 µg/µl in anhydrous DMSO. The aa-modified RNA

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1165

polynucleotides were resuspended in a final volume of 90 ul of 100 mM $Na_2B_4O_7$ pH 8.5, and heated to 37°C for 20 min. 110 µl of the resuspended dyes were individually mixed with 90 µl of an aa-modified RNA polynucleotide. The mixture was incubated in the dark at room temperature for 30 min-2 hr.

[00255] The dye-colored, aa-modified RNA polynucleotides were purified from the mixture using an RNeasy kit (Qiagen®) following the manufacturer's instructions. Equal amounts of each of the 24 dye-colored RNA polynucleotides were mixed for a final concentration of 40 nM of each polynucleotide ("color segment mix"). This mixture may be combined with a population of nanoreporter backbones to generate a population of unique nanoreporters, as described in Example 3 below.

Example 3: Generating a Labeled DV1 Nanoreporter Library

[00256] To produce a library of labeled DV1 Nanoreporter molecules, each code-specific DNA backbone from Example 1 was ligated separately to a probe region, *i.e.*, target-specific nucleotide sequence, for a specific gene of interest. The probe was ligated to the backbone via an oligonucleotide, which serves as a bridge between the backbone and the specific probe. Specifically, a master mix containing a universal oligonucleotide that served as a ligation "bridge" plus ligase buffer was added to individual wells of 96-well plates containing normalized (10µM) target-specific oligonucleotide probes (35-50 bases). After a short incubation at 37°C to anneal the probe oligonucleotide to the complementary portion of the bridge oligonucleotide, ligation was initialized by addition of 1.2 pmoles of individual single-stranded nanoreporter backbone per well, additional ligation buffer, and T4 ligase. Plates were incubated at 37°C in a 96-well thermocycler for 2h. Ligation reactions were desalted via centrifugation through G-50 Sephadex columns in a 96-well format. The individual ligated backbones were pooled, ethanol-precipitate, and resuspended in 10 mM Tris pH 8.0 at 20 nM.

[00257] To produce dye-labeled nanoreporters, the following was mixed: 2 ml 20X SSPE pH 6.5, 28.05 ml H2O, 5 ml pooled DV1 nanoreporter backbones from above, 1.2 ml ethanol, and 3.75 ml colored segment mix from Example 2. The mixture was incubated at 75°C for 2 hours. To remove excess segments, the mixture was purified over an oligonucleotide column coupled to the reverse complement of the G-repeat sequence common to all of the backbone sequences (*see* SEQ ID NO:26 for the reverse complement sequence). To produce unlabeled "capture" probes, the nanoreporter backbones were biotinylated near or at the 3' region.

Example 4: Comparison of DV1 and M13 Nanoreporter Systems

[00258] A 40-gene DV1 reporter library was made as described in Examples 1 to 3, and a comparable M13 library was also made, as described in U.S. Application No 12/100,990, herein incorporated by reference. The DV1 and M13 libraries contained identical probe (*i.e.*, target-specific nucleotide)

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1166

sequences and were used to assay the same samples – 36 cell lysates, assayed in triplicate. Standard hybridization, purification and imaging protocols were followed, as described in Geiss *et al.*, *Nature Biotechnology* 26:317-324, 2008, herein incorporated by reference in its entirety. Generally, nanoreporter libraries were hybridized to RNA-containing cell lysates, excess reporters were removed by washing, reporters were bound to a surface upon which the reporters were immobilized and stretched, the surface was imaged, and the images were analyzed.

Hybridization Reactions

[00259] Detection of cellular transcripts was carried out in multiplexed hybridization reactions, which utilized a dual nanoreporter system having both labeled nanoreporter probes and unlabeled nanoreporter probes, *i.e.*, "capture" probes. Each sample was hybridized in triplicate with final concentrations of the hybridization reagents as follows: 200pM each unlabeled, biotinylated capture probe (capture probe), 40pM each labeled reporter probe, 5X SSPE (pH 7.5), 5X Denhardt's reagent (Sigma), 100ng/µl sheared salmon sperm DNA (Sigma), and 0.1% Tween-20. Each 30µl hybridization reaction also contained 100ng total RNA at a final concentration of 3.3ng/µl. Reagents were mixed and incubated at 65°C in a thermocycler block with a heated lid for 20 hours.

Post-hybridization purification

[00260] To remove unhybridized reporters, reactions were purified over magnetic beads (Invitrogen™) coupled to oligonucleotides complementary to the 3'-repeat sequence contained on every capture probe. Reactions were first diluted to 1X SSPE in 0.1% Tween-20/TE and allowed to bind to beads at 22.5°C for 30 minutes with continuous rotation. The beads were washed three times in 150µl of 0.1X SSPE/0.1% Tween-20 and the hybridized complexes eluted in 100µl of 0.1X SSPE/0.1% Tween-20 for 15 minutes at 45°C. After elution, samples were purified a second time to remove excess capture probes by binding to magnetic beads coupled to oligonucleotides complementary to the 5'-repeat sequence contained on every reporter probe. The elutions from the anti-3'-repeat beads were brought to a final concentration of 1X SSPE by addition of 50µl of 3X SSPE/0.1% Tween-20 and bound for 15 minutes at 22.5°C with rotation. Beads were washed as above and eluted in 30µl of 0.1X SSPE/0.1% Tween-20 at 45°C. The doubly-purified samples were then prepared for capture as described below.

NanoString reporter capture, stretching, and imaging

[00261] One microliter of 1/5000 dilution of a 0.1% solids solution of a custom-formulation of Tetraspeck fluorescent microspheres (Invitrogen™) was added to each sample. Samples were loaded into a NanoString fluidic device made by lamination of laser-machined cast acrylic with a coverslip coated with streptavidin (Optichem®, Accelr8 Technology Corporation) using a laser-cut double-sided adhesive layer (Fralock) to generate 30i.tm deep microfluidic channels. The samples were driven through the channel by hydrostatic pressure and bound specifically by the biotinylated 3' end of the capture probe. After capture, the surface was washed once with 90µl of 1X TAE and prepared for stretching by the addition of 40µl of TAE to each well. Reporter probes were stretched and aligned by

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1167

applying 160V/cm for 1 minute along the fluidic channel. Stretched reporters were then immobilized to the surface by addition of 60µl of a 500nM solution of a biotinylated oligonucleotide complementary to the 5'-repeats present on the 5' end of all reporter probes. The current remained on for 5 minutes, throughout the immobilization process. After immobilization, the TAE solution was removed and replaced with a custom formulation of the anti-photobleaching reagent SlowFade (Invitrogen) for imaging.

[00262] Slides were imaged on a Nikon Eclipse TE2000E equipped with Perfect Focus, a 1.4 NA Plan Apo VC 60X oil-immersion lens (Nikon), an X-cite 120 metal halide light source (Exfo Corporation), an automated H117 stage (Prior Scientific), and a SmartShutter (Sutter Instrument). For each field of view, 4 images at different excitation wavelengths (480, 545, 580 and 622) were acquired with an Orca Ag CCD camera (Hamamatsu) under control of either Metamorph (Universal Imaging Corporation) or custom software.

Image processing

[00263] Image processing was performed on 4 images (one for each wavelength) on a FOV-by-FOV basis. The custom algorithm treats each FOV as a fundamental block in which the following basic steps are performed: 1) spot identification, 2) image registration, 3) spatial clustering to produce strings, and 4) string classification.

[00264] In the first step of the algorithm, spots were identified. The background intensity level of each channel was computed and used to threshold the image into signal and background, where signal regions are the result of a specific wavelength of light observed as a point spread function (PSF). The signal mask was segmented using a custom Watershed algorithm. The segmented regions were then labeled, parameterized, and filtered to remove non-PSF spots. The remaining spots were centrally archived for use in registration and reporter calling.

[00265] Image registration was performed on each FOV based on archived spots that correspond to fluorescent beads (fiducials) that were bound to the imaging surface (see NanoString reporter capture, stretching, and imaging). The archived spots were cross-referenced to identify inter-channel clusters of spots that meet fiducial requirements (interchannel intensity thresholds and ratios). Clusters that met requirements were archived as fiducials. The final list of fiducials represented the spatial transforms that occurred between channels during image acquisition. Spatial offsets were as large as 5-6 pixels. The spatial transform was solved for using the observed fiducial centroids and their pre transform (assumed) coincident centroids ($X_2 = X_1 * T$). The inverse transform was then applied to all identified spots to restore their original centroids.

[00266] After spot identification and image registration, spots were assembled into "strings" via clustering. At this point, each string was filtered to remove any spots attributed to bleed-though signal. The filtered strings were then classified as reporters or non-reporters. To be classified as a reporter the string must contain the correct number of spots, meet specific spot-to-spot spacing

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1168

thresholds (1.2-2.9 pixels), and meet acceptable linearity and orientation requirements. Clusters that were classified as reporters were then counted and summed for each gene over all FOVs.

NanoString data normalization and analysis

[00267] To account for slight differences in hybridization and purification efficiency, data was normalized to the average counts for all control spikes in each sample. To determine if a gene was "detected" by the NanoString system, the triplicate measurements obtained for each experimental gene were compared to triplicate measurements for the negative controls. For a gene to be categorized as detected, the average counts for the experimental gene had to be greater than the average counts for the 2 negative controls, and the Student's T-test P-value had to be less than 0.05.

[00268] Results are shown for a representative three of the 40 genes probed, comparing the DV1 system to the M13 system. Figure 5 shows the results for IL-8 (results for GAPDH, and TSC22D3 are not shown). The overall profile of the expression of each gene is the same for both systems (*i.e.*, greater or lesser relative expression between samples), but the number of counts for each assay is >100-fold higher using DV1 (note the difference in scale for the M13 and DV1 graphs).

[00269] Results of a comparison of the expression of all 40 genes between the systems are also shown in two representative samples (*see* Figure 6). In Figure 6, the R^2 value shows an approximately 90% correlation between the relative expression of all the genes in the two systems, while the slope of the line indicates a greater than 100-fold increase in the actual number of counts using the DV1 reporters as compared to the M13 reporters.

Example 5: Comparison of DV1 and M13 Nanoreporter Systems

[00270] A 148-probe DV1 reporter library was made as described in Examples 1 to 3, and a comparable M13 library was also made, as described in U.S. Application No 12/100,990, herein incorporated by reference. The DV1 and M13 libraries contained identical probe (*i.e.*, target-specific nucleotide) sequences and were used to assay the same samples – 26 mouse RNA samples. Standard hybridization, purification and imaging protocols were followed, as described in Example 4.

[00271] The results for the detection of GusB expression are shown in Figures 7, respectively. Results for the detection of Mod1, Acot4, and Atg7 expression are not shown. The measurement of relative expression levels of a given gene across the various samples was comparable between the two reporter types. However, an average 6-fold increase was seen for the absolute number of counts measured by the DV1 reporters compared to the M13 reporters. Figure 8 provides a graph showing the "valid reporters," which refers to strings of spots that can be interpreted by the image analysis software as a "countable" reporter, described as a percentage of the total number of binding events in a field-of-view. A significantly higher percentage of DV1 reporters are countable, in this experiment leading to a 3-fold gain in data (from 12.5% to 38%).

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1169

[00272] A correlation of the expression of each gene in one sample as measured by M13 and DV1 was determined. The results show a 77% correlation between the two systems and an average 6-fold increase with DV1 (data not shown).

Example 6: Optimization of Dye Incorporation into Complementary RNA Polynucleotides

[00273] To determine the optimum fluorescent dye concentration in a given nanoreporter molecule, fluorescent dyes were incorporated into the complementary RNA polynucleotide sequences in various amounts. Nanoreporter molecules having various amounts of dye incorporation were then tested for signal intensity in a standard nanoreporter detection assay.

[00274] Specifically, to determine the optimum degree of nucleotide spacing between each incorporated dye molecule, fluorescent dye molecules were incorporated into complementary RNA polynucleotides at every into every 8, 10, 12, 16, or 24 nucleotide bases. This was accomplished by generating novel sequences which contained a regular pattern of T's spaced every 8, 10, 12, 14, 16 and 24 bases, as shown in Figures 9-12. These sequences were approximately 1100 bases long and were assembled into a single 6600 base reporter backbone plasmid, which was converted into a single-stranded backbone. Each template was transcribed to produce the reverse complementary RNA, and aliquots of each RNA were coupled separately to four colors: blue (Alexa 488), green (Cy3(Amersham)), yellow (Alexa 594), and red (Alexa 647). A series of labeled reporters using the various colors were generated by annealing labeled RNA segments to the backbone. These reporters were imaged as described herein and the brightness of the spot resulting from the dye-coupling at each spacing in each color was determined (data not shown). Spot intensity measurements are in arbitrary units.

[00275] These experiments show that the brightness of the spots can be manipulated by designing the underlying sequence to contain greater or fewer regularly repeated bases. Here, a spacing of 8 nucleotides gives the brightest spots. Spacings of fewer than 8 were not tested due to predictions of stearic interference of neighboring dye molecules at closer spacings; however, it is likely that some spots could be made even brighter by closer spacings. The closest possible spacing will vary from dye to dye, but 8 bases provide a spacing that should accommodate all of the commercially suitable dyes. The bases do not need to be regularly spaced.

Example 7: Considerations Regarding Nanoreporter Hybridization Kinetics

BACKGROUND

[00276] Solution hybridizations with a large excess of probe over target follow pseudo-first order kinetics. In this regime the speed of the reaction depends only on the probe concentration and not on the target concentration. For a two-probe, one-target strategy to provide accurate information on the concentration of a target in solution, the probes should both be present in excess of the target. The possible concentration range is preferably therefore bounded on the lower end by the concentration of

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1170

the target. However, the useful concentration range for the nanoreporter technology described herein is practically bounded on the lower end by the amount of time needed to perform the hybridization.

HYBRIDIZATION KINETICS

[00277] In preferred embodiments, target detection and quantification assays are performed in which the target (T) must hybridize to both a reporter probe (R) and a capture probe (G) to be detected (for example by affinity selection and detection of complexes comprising only (R) and (G), which in turn only form complexes in the presence of (T)). Assuming that these reactions are irreversible, there are four possible elementary reactions that occur.

$$R + T \xrightarrow{k_1} RT$$

$$T + G \xrightarrow{k_2} TG$$

$$RT + G \xrightarrow{k_3} RTG$$

$$R + TG \xrightarrow{k4} RTG\,.$$

[00278] Because RT and TG are intermediate complexes of two out of the three species, these four reactions can be simplified to $R + T + G \rightarrow RTG$.

[00279] However, to quantitatively calculate the rate of production of RTG (the reporter-target-capture probe complex), all four reactions must be considered. The differential equations describing the system are:

$$\frac{dC_G}{dt} = -k_2 C_G C_T - k_3 C_G C_{RT} \qquad \frac{dC_G}{dt} = -k_1 C_R C_T - k_4 C_R C_{TG}$$

$$\frac{dC_T}{dt} = -k_2 C_G C_T - k_3 C_R C_T$$

$$\frac{dC_{TG}}{dt} = k_2 C_G C_T - k_4 C_R C_{TG} \qquad \frac{dC_{RT}}{dt} = k_1 C_R C_T - k_3 C_G C_{RT}$$

$$\frac{dC_{RTG}}{dt} = k_4 C_R C_{TG} + k_3 C_G C_{RT}$$

where C_R, C_T, C_G, C_{RT}, C_{TG}, and C_{RTG} are the concentrations of the various species, and 1(1-1(4 are the kinetic constants for the four elementary reactions. Values for these kinetic constants when the probes and targets are complementary single-stranded molecules (*i.e.*, when there is no purification tag on the capture probe and no reporter) can be calculated from data available in the literature (Wetmur, *J. Annu. Rev. Biophys. Bioeng.* 1976.5:337-361).

$$k = k_N \frac{\sqrt{L}}{N} \frac{\alpha_{salt}}{\alpha_{ref}}$$

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1171

[00280] In the above equation, k_N is the nucleation rate constant, L is the nucleic acid length (in base pairs), N is the nucleic acid complexity (equal to L for non-repetitive sequences) and a_{salt} and a_{ref} are corrections for salt concentration (Britten *et al.*, 1974, *Methods in Enzymology* 29E:363-406). In the nanoreporter systems described herein, the kinetic constants will depend on the sizes of the attached capture probe tags and reporter probe. Without being bound by any theory, it is the inventors' belief that the kinetic constants will have the same dependence on length that an elementary reaction has on the diffusion constants of the reactants.

$$k = k_N \frac{\sqrt{L}}{N} \frac{\alpha_{salt}}{\alpha_{ref}} \frac{D_1 + D_2}{2D_{50}}$$

[00281] In the above equation D1 and D2 are the diffusion constants of the two reacting species (*see* the reactions above) and D_{50} is the diffusion constant of a 50-mer single-stranded DNA molecule. Assuming a 100-base single-stranded target, 100-base single-stranded capture probe, and 7200-base double stranded reporter, the relevant kinetic constants are

$k_1 = 2.64 \times 10^5$ L/mol/s

$k_2 = 6.55 \times 10^5$ L/mol/s

$k_3 = 3.99 \times 10^5$ L/mol/s

$k_4 = 1.91 \times 10^5$ L/mol/s

[00282] Numerically solving the system of differential equations with these kinetic constants (assuming at least a 10-fold excess of probes over target) yields the prediction that 5 pM reporter and 5 pM capture probe will drive hybridization to 10% of completion in an overnight reaction (16-18 hours). At concentrations lower than 5 pM, the amount of completely hybridized molecules is likely impractical to measure. Thus, in a preferred embodiment, the lower concentration of a nanoreporter component (capture probe and/or reporter probe) is 5 pM.

ENTANGLEMENT OF REPORTERS

[00283] As probe concentrations increase, theory predicts that hybridization kinetics speed up without bound - the only limit being the solubility of the probes. However, the reporter probe can be very large compared to the target-specific sequence in the nanoreporter systems of the invention. Without being bound by any theory, the inventors believe that by its attachment to the reporter probe the kinetics of the target-specific sequence are altered from classical solution hybridization kinetics. Because the reporter probe is a large, polymeric molecule, it can have long-lived interactions (entanglements) with other nanoreporters when they come into contact. At low concentration the probability of two polymers becoming entangled is small, but as the concentration and/or size of a polymer in solution increases, these interactions become more and more common. In the extreme case of very long molecules at very high concentration the polymers form a permanent network, or gel, in solution. For solution hybridization to occur, a probe (*e.g.*, a nanoreporter probe)/target pair must diffuse through solution until they contact one another and a hybridization nucleus forms. Classically,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1172

hybridization reactions are not diffusion limited because the translational diffusion of the molecules is faster than the nucleation of the hybridization (*i.e.*, the probe and target diffuse together and interact many times before a nucleation occurs). In dilute solution its large size will slow the translational diffusion of the reporter probe, but may not significantly affect the kinetics. At some intermediate concentration, the reporter probes take up almost all of the space in the solution, effectively forming a permanently entangled gel, and can no longer diffuse in solution. However, the capture probe and the targets are smaller molecules that are believed to still diffuse through the entangled reporter probes, allowing hybridization to take place (although possibly at a slower rate). The inventors also believe that at some higher concentration the reporter probe in solution will also hinder the movement of the capture probe and the targets to the point that the reaction becomes diffusion limited. This concentration (which is not quantitatively known and depends upon the reporter probe structure, the capture probe structure, and the target size) is the upper limit of the useful concentration range in the nanoreporter system, and can be empirically determined by one of skill in the art guided by the principles described herein.

LENGTH DEPENDENCE OF KINETICS

[00284] Since the limiting upper concentration for hybridization depends upon both the reporter structure and capture probe structure (of which there are many possible variations), a theoretical framework to predict the permutations of useful concentration ranges is useful in the practice of the invention. Classical theory predicts that hybridization kinetics depend only on the size of the smaller probe. Theory would therefore predict that the size of the reporter will not play a role in the hybridization kinetics as long as both the target molecule and the capture probe are significantly smaller. Theory then predicts that the rate of hybridization (for a constant target length) depends on $1/L^{1/2}$, where L is the length of the capture probe, due to steric inhibition of hybridization. Consequently, the kinetics of hybridization will be faster with smaller capture probes. As the capture probe length increases, the hybridization rate should decrease as $1/L^{1/2}$. If a constant capture probe length is assumed, then the range of reporter lengths and concentrations that will result in a measurable mount of hybridization events can be defined. Once a reporter size has been defined, then the approximate range of capture probe sizes can be determined. This is an iterative process, but may give good starting points from which to gather data to generate detailed empirical guidelines, given that the theories that the inventors' rationale is based upon were generated from hybridization data in systems that do not employ a reporter probe.

ENTANGLEMENT THRESHOLD

[00285] A reporter probe is essentially a polymer in free solution, which behaves as a random coil. The volume occupied by a single reporter, V_p, can be calculated from polymer physics theories according to the Freely-Jointed Chain model (FJC, for a flexible polymer, such as single-stranded DNA or RNA) or the Worm-Like Chain model (WLC, for a stiff polymer such as double-stranded DNA or a reporter). For either model

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1173

$$V_p = \frac{4}{3}\pi R_g^3$$

where R_g is the radius of gyration. For the FJC

$$R_g = b\left(\frac{N}{6}\right)^{0.6}$$

where b is the segment length and N is the number of segments in the chain. For the WLC

$$R_g = \sqrt{\frac{1}{6}Nb^2 - \frac{b^2}{4} + \frac{b^2}{4N}\left(1 + \frac{1}{2N}\left(e^{-2N} - 1\right)\right)}$$

[00286] The entanglement threshold concentration is defined as concentration is defined as the concentration where the entire volume of the solution is occupied by the reporters.

$$C^* = \frac{3}{4\pi R_g^3 N_A}$$

where N_A is Avogadro's number. Above this concentration it is assumed that the translational diffusion of the reporters is severely restricted. The entanglement threshold concentration varies with the reporter structure. As the reporter length increases, the entanglement threshold decreases (as $1/L^{1.5}$). From the equations above, the theoretical entanglement threshold for reporter probes with different spot sizes and different lengths can be calculated. The result of such calculations shows that for a 7200 by RNA/DNA hybrid reporter probe with 8 label attachment regions of about 900 by each, the entanglement threshold is about 70 nM.

[00287] If both the target and the capture probe are much smaller than the reporters, then they will most likely be free to diffuse through the solution even at these high concentrations of reporters. Initial data indicates that hybridization kinetics do not slow appreciably up to a concentration of 80 nM with a 7200-bp reporter probe, a 100-base target, and a 100-base capture probe.

EFFECT OF ENTANGLEMENT THRESHOLD ON MULTIPLEXING

[00288] Assuming that the maximum concentration for reporters in a hybridization reaction is C*, then the concentration of each reporter (specific to a particular target) is equal to C*/M, where M is the multiplex of the reaction (number of different targets being addressed simultaneously). Conversely, the possible multiplex level for a particular reporter structure can be calculated from the lower limit of probe concentration (C_p from kinetics ~ 10 nM) and the entanglement threshold

$$M = \frac{C^*}{C_p}$$

[00289] If the number of nanoreporter codes available does not depend on reporter probe size, then the multiplexing of the nanoreporter depends primarily on the reporter probe size and concentration (since it is much larger than the capture probe). Because the capture probe makes an insignificant contribution to entanglement during hybridization, it is the inventors' belief that the concentration of the capture probe can be increased far above the concentration of the reporter probe. In Table 4 below,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1174

the maximum total capture probe concentration ([G]) is set to 1000 nM for all reporter concentrations. This difference in concentration of capture probe and reporter probe is an adjustable parameter. Preliminary experiments show that in a multiplex hybridization reaction with a 7200 by reporter and 100b capture, 40pM of each reporter probe and 200 pM of each capture probe results in near complete hybridization in an overnight reaction.

OPTIMAL SIZE AND CONCENTRATION RANGES

[00290] Below in Table 4 is a summary of the optimal useful size and concentration ranges of the capture probe and reporter probe at different multiplexing as approximated by the above theories. It is the inventors' belief that capture probes up to about 200 bases will be practical for most applications

Table 4: Optimal size and concentration ranges of reporter probe, capture probe and target, as well as multiplicity of probes, in the nanoreporter systems of the invention.

Reporter Length (bp)	Capture Length (b)	Minimum [R] (pM)	Minimum [G] (pM)	Maximum [R](nM)	Maximum [G](nM)	Max Multiplex
2000	100	5	5	603	1000	114417
2000	50	4	4	603	1000	161811
2000	200	7	7	603	1000	80905
3000	100	6	6	292	1000	45182
3000	50	5	5	292	1000	63897
3000	200	9	9	292	1000	31948
4000	100	7	7	178	1000	23912
4000	50	5	5	178	1000	33817
4000	200	11	11	178	1000	16908
5000	100	8	8	123	1000	14746
5000	50	6	6	123	1000	20854
5000	200	12	12	123	1000	10427
6000	100	9	9	91	1000	9988
6000	50	6	6	91	1000	14125
6000	200	13	13	91	1000	7062
7200	100	10	10	68	1000	6792
7200	50	7	10	68	1000	6792
7200	200	14	10	68	1000	6792
8000	100	11	11	57	1000	5444
8000	50	7	7	57	1000	7699
8000	200	15	15	57	1000	3850
10000	100	12	12	40	1000	3419
10000	50	8	8	40	1000	4835
10000	200	17	17	40	1000	2417

[00291] From the foregoing it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1175

deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.

[00292] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1176

CLAIMS

1. A population of uniquely labeled nanoreporter probes, wherein the nanoreporter probe comprises:

i) an unique target-specific region; and

ii) a region comprising an unique, designed nanoreporter wherein said nanoreporter comprises a single-stranded nucleic acid backbone, said backbone comprising a plurality of label attachment regions covalently attached together in an unique linear combination, wherein the label attachment regions are selected from a population of designed polynucleotide sequences, wherein the polynucleotide sequence is hybridized to a complementary polynucleotide sequence having attached thereto one or more detectable molecules; wherein each complementary polynucleotide sequence has designated a specific detectable molecule; and wherein each nanoreporter has a detectable signal that distinguishes it from other nanoreporters in said population.

2. The population of claim 1 wherein each label attachment region of each backbone is different from the other label attachment regions in that same backbone.

3. The population of claim 1 wherein said label attachment region comprises about 800 to 1300 nucleotide bases and has a G/C content of about 50% and wherein said complementary polynucleotide sequence has a G/C ratio of about 1/1.

4. The population of claim 3 wherein said complementary polynucleotide sequence has a G/C ratio of about 3/2.

5. The population of claim 1 wherein said label attachment regions are generated from the template sequences selected from the group consisting of SEQ ID NOS: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, and 24.

6. The population of claim 1 wherein said nanoreporter probe further comprises a constant region, wherein the constant region comprises a plurality of repeat nucleotide sequences.

7. The population of claim 1 wherein the label attachment regions comprise a similar adenine content.

8. The population of claim 7 wherein the adenine bases are spaced at least an average of every 8 to 16 nucleotide bases.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1177

9. The population of claim 1 wherein the label attachment regions comprise a regularly repeated pattern of adenine bases.

10. The population of claim 9 wherein the adenine bases are spaced about every 8 to 16 nucleotide bases.

11. The population of claim 1 wherein the label attachment regions comprise a thymine content of about 35-45%.

12. The population of claim 1 wherein the complementary polynucleotide sequence comprises a RNA polynucleotide sequence.

13. The population of claim 12 wherein the RNA polynucleotide sequence comprises at least one amino-allyl modified uracil base.

14. The population of claim 13 wherein the detectable molecule is attached to the amino-allyl modified uracil base.

15. The population of claim 12 wherein the RNA polynucleotide sequence comprises a plurality of amino-allyl modified uracil bases that are spaced at about an average of every 8 to 16 bases in said RNA polynucleotide sequence.

16. The population of claim 15 wherein the detectable molecule is attached to each of the allyl modified uracil bases.

17. The population of claim 1 wherein the detectable molecules are fluorescent dyes.

18. The population of claim 1, wherein when said population of nanoreporter probes is used to detect a population of target molecules in a sample by a method comprising (a) forming at least one molecular complex with each nanoreporter probe comprising at least one target molecule and at least one nanoreporter probe and (b) individually counting the presence of one or more molecules of each of said molecular complex in said sample, the numbers of counts above background of each of said molecular complex after normalization of the sample is at least two fold higher when compared to a nanoreporter probe comprising M13 DNA.

19. The population of claim 1 where the melting temperature (Tm) of said complementary polynucleotide sequences when hybridized to its label attachment regions is about 80°C or higher.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1178

20. A method for determining the presence of at least one target molecule in a sample, comprising:

(1) forming at least one molecular complex comprising:

(a) at least one target molecule, and

(b) at least one probe comprising an unique target-specific region and a region comprising an unique, designed nanoreporter, wherein said nanoreporter comprises a single-stranded nucleic acid backbone, said backbone comprising a plurality of label attachment regions covalently attached together in an unique linear combination, wherein each label attachment region is hybridized to a complementary polynucleotide sequence having attached thereto one or more detectable molecules, wherein each complementary polynucleotide sequence has designated a specific detectable molecule; and

(2) individually counting the presence of one or more molecular complex or at least part of said at least one molecular complex to determine the presence of said at least one target molecule in the sample.

21. The method of claim 20 wherein the percentage of valid molecular counts of said molecular complex is higher that about 12.5%.

22. The method of claim 20 wherein the numbers of counts is at least two fold higher than the counts obtain when using a nanoreporter probe comprising M13 DNA.

23. The method of claim 22 wherein said nanoreporter comprising M13 DNA comprises a single–stranded backbone comprising a plurality of M13 DNA regions covalently attached together wherein each region is hybridized to a complementary polynucleotide sequence having attached thereto one or more detectable molecules.

24. The method of claim 23 wherein the Tm of said complementary polynucleotide sequences when hybridized to its label attachment regions is higher than the melting temperature of polynucleotide sequences complementary to M13 DNA when hybridized to said nanoreporter probe comprising M13 DNA.

25. The method of claim 20 wherein the Tm of said complementary polynucleotide sequences when hybridized to its label attachment regions is about 80°C or higher.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1179

26. The method of claim 20 wherein each label attachment region comprises about 800 to 1300 nucleotide bases and has a G/C content of about 50% and the complementary polynucleotide sequence has a G/C ratio of at least 1/1.

27. The method of claim 20 wherein each selected label attachment region is different from the other selected label attachment regions in the backbone.

28. The method of claim 20 wherein each label attachment region is generated from the template sequences selected from the group consisting of SEQ ID NOS: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, and 24.

29. The method of claim 20 wherein said probe further comprises a constant region, wherein the constant region comprises a plurality of repeat nucleotide sequences.

30. The population of claim 20 wherein the label attachment regions comprise a similar adenine content.

31. The population of claim 30 wherein the adenine bases are spaced at least an average of every 8 to16 nucleotide bases.

32. The method of claim 20 wherein the label attachment regions comprise a regularly repeated pattern of adenine bases.

33. The method of claim 32 wherein the adenine bases are spaced about every 8 to 16 nucleotide bases.

34. The method of claim 20 wherein the label attachment regions comprise a thymine content of about 35 to 45%.

35. The method of claim 20 wherein the complementary polynucleotide sequence comprises a RNA polynucleotide sequence.

36. The method of claim 35 wherein the RNA polynucleotide sequence comprises at least one amino-allyl modified uracil base.

37. The method of claim 32 wherein the detectable molecule is attached to the amino-allyl modified uracil base.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1180

38. The method of claim 36 wherein the RNA polynucleotide sequence comprises a plurality of amino-allyl modified uracil bases that are spaced at least an average of every 8 bases in said RNA polynucleotide sequence.

39. The method of claim 37 wherein the detectable molecule is attached to each of the allyl modified uracil bases.

40. The method of claim 20 wherein the detectable molecules are fluorescent dyes.

41. The method of claim 20 further comprising determining the presence of a plurality of target molecules by a method comprising forming a plurality molecular complexes, each complex comprising (a) at least one target molecule and (b) at least one probe comprising an unique target-specific region and a region comprising an unique, designed nanoreporter, wherein said nanoreporter comprises a single-stranded nucleic acid backbone, said backbone comprising a plurality of label attachment regions covalently attached together in an unique linear combination, wherein each label attachment region is hybridized to a complementary polynucleotide sequence having attached thereto one or more detectable molecules, and wherein each probe comprises a different nanoreporter region.

42. The method of claim 41 wherein the presence of at least 5, 10, 20, 30, 50, 100, 200, 300, or 500 different target molecules is determined.

43. The method of claim 20 wherein said target molecule is a nucleic acid.

44. The method of claim 43 wherein said nucleic acid comprises at least one heritable mutation, at least one somatic mutation, at least one single nucleotide polymorphism (SNP), at least one point mutation, at least one deletion mutation, at least one insertion mutation, at least one chromosomal translocation, or combinations thereof.

45. The method of claim 20 wherein said target molecule is a diagnostic indicator.

46. An uniquely labeled nanoreporter probe comprising:

i) an unique target-specific region; and

ii) a region comprising a plurality of designed label attachment regions covalently attached together in a linear combination, wherein each label attachment region comprises about 800 to 1300 nucleotide bases and has a G/C content of about 50%, wherein each selected label attachment region is different from the other selected label attachment regions, and

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1181

wherein each label attachment region is hybridized to a complementary polynucleotide sequence having attached thereto one or more detectable molecules, wherein the complementary polynucleotide sequence has a G/C ratio of at least 1/1.

47. The method of claim 46 wherein the complementary polynucleotide sequence has a G/C ratio of about 3/2.

48. The nanoreporter probe of claim 46 further comprising a constant region, wherein the constant region comprises a plurality of repeat nucleotide sequences.

49. The nanoreporter probe of claim 46 wherein the label attachment regions comprise a regularly repeated pattern of adenine bases.

50. The nanoreporter probe of claim 49 wherein the adenine bases are spaced about an average of every 8 to 16 nucleotide bases.

51. The nanoreporter probe of claim 46 wherein the label attachment regions comprise a thymine content of about 35-45%.

52. The nanoreporter probe of claim 46 wherein the complementary polynucleotide sequence comprises a RNA polynucleotide sequence.

53. The nanoreporter probe of claim 52 wherein the RNA polynucleotide sequence comprises at least one amino-allyl modified uracil base.

54. The nanoreporter probe of claim 53 wherein the detectable molecule is attached to the amino-allyl modified uracil base.

55. The nanoreporter probe of claim 52 wherein the RNA polynucleotide sequence comprises a plurality of amino-allyl modified uracil bases that are spaced at about an average of every 8 to 16 bases in said RNA polynucleotide sequence.

56. The nanoreporter probe of claim 55 wherein the detectable molecule is attached to each of the allyl modified uracil bases.

57. The nanoreporter probe of claim 46 wherein the detectable molecules are fluorescent dyes.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1182

58. The nanoreporter probe of claim 46 wherein said plurality of label attachment regions are generated from the template sequences selected from the group consisting of SEQ ID NOS: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, and 24.

59. A method of preparing at least one uniquely labeled nanoreporter comprising:

i) combining a plurality of label attachment region each comprising about 800 to 1300 nucleotide bases and a G/C content of about 50%, wherein each selected label attachment region is different from the other selected label attachment regions;

ii) covalently attaching the plurality of label attachment regions to each other in linear combination; and

iii) hybridizing a complementary polynucleotide sequence to said label attachment region, wherein said complementary polynucleotide sequence is having attached thereto one or more detectable molecules.

60. The method of claim 59 further comprising preparing a labeled nanoreporter probe by attaching the labeled nanoreporter to a target specific region.

61. The method of claim 59 wherein said complementary polynucleotide sequence has a G/C ratio of at least 1/1.

62. A kit for preparing at least one uniquely labeled nanoreporter comprising:

a) at least three label attachment regions each comprising about 800 to 1300 nucleotide bases, a G/C content of about 50%; and

b) at least three complementary polynucleotide sequences having attached thereto a detectable molecule, wherein the complementary polynucleotide sequences has a G/C ratio of at least 1/1.

63. The kit of claim 62 further comprising at least three target specific probes.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1183

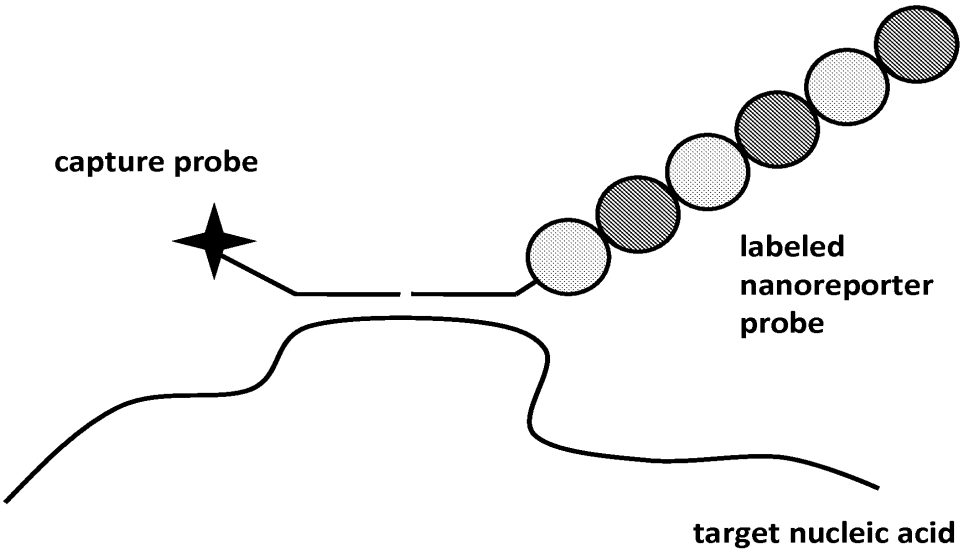

capture probe
labeled
nanoreporter
probe
target nucleic acid

Figure 1

Figure 2A

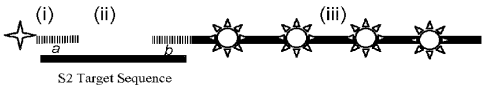
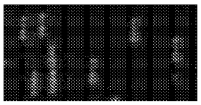


Figure 2E



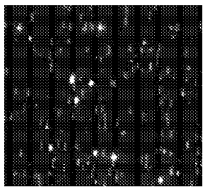


Figure 2B

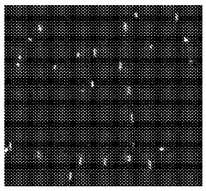


Figure 2C

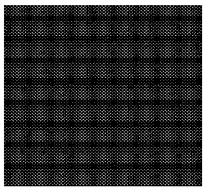


Figure 2D

Figure 2

Figure 3A

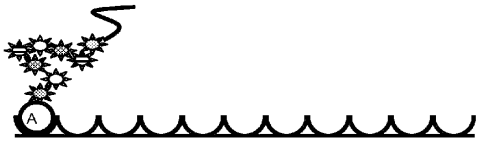


Figure 3B

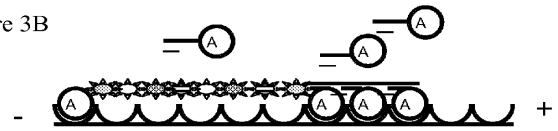


Figure 3

Figure 4A

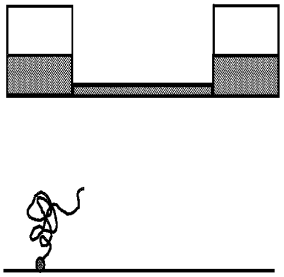


Figure 4B

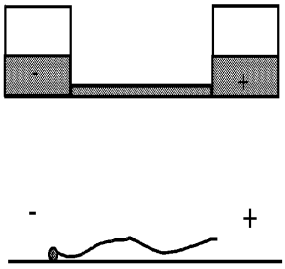


Figure 4C

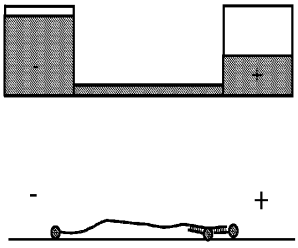


Figure 4

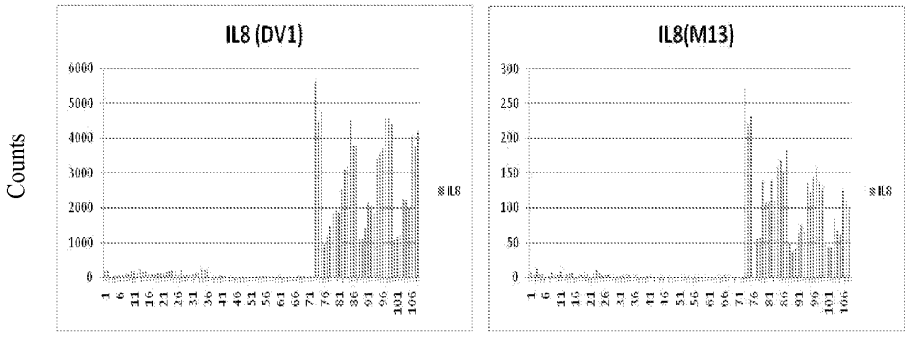
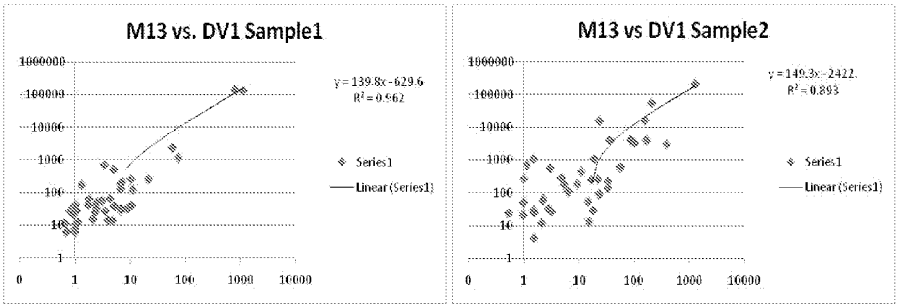

IL8 (DV1)
Counts
6000
5000
4000
3000
2000
1000
0
IL8
IL8(M13)
300
250
200
150
100
50
0
IL8

Figure 5

40 reporters from 2 'representative' samples:



89-96% correlation between the two systems, with greater than 100-fold increase in signal with DV1

Figure 6

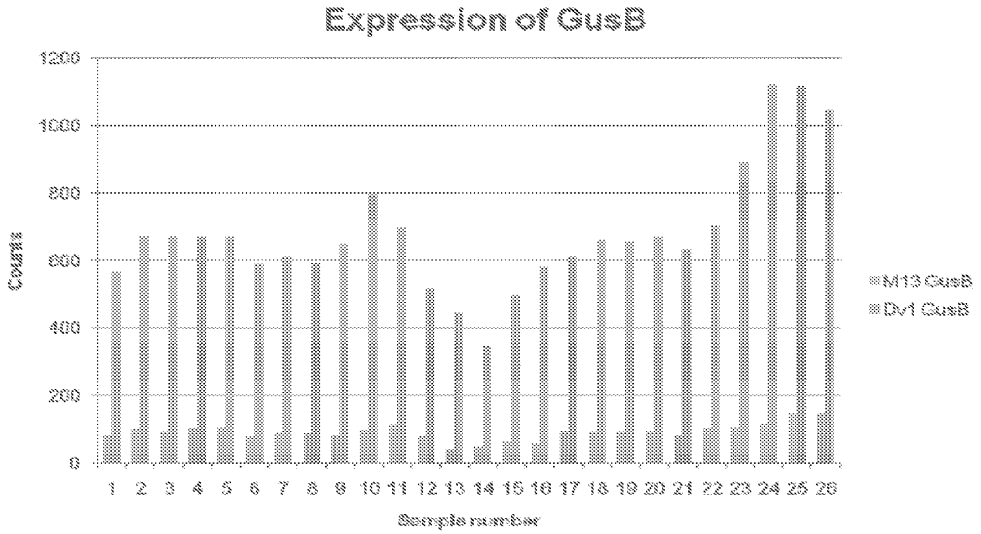

Expression of GusB
1200
1000
800
600
400
200
0
Counts
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
Sample number
M13 GusB
Dv1 GusB

Figure 7

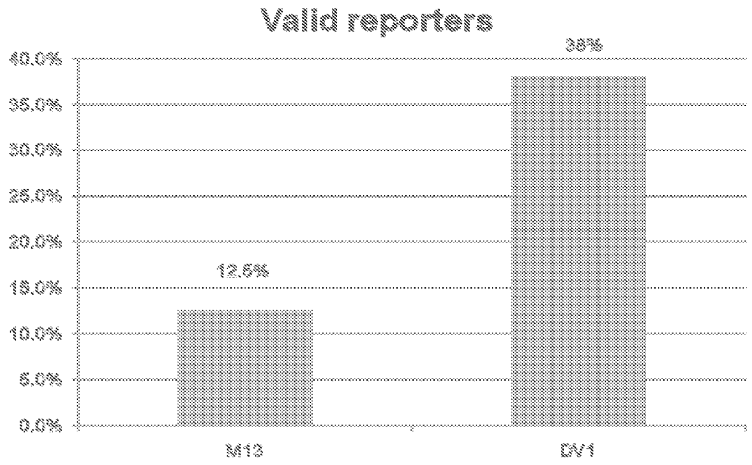

Valid reporters
40.0%
35.0%
30.0%
25.0%
20.0%
15.0%
10.0%
5.0%
0.0%
12.6%
38%
M13
DV1

Figure 8

AAGCTTGGCCACACAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGT
TTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCT
GGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACC
TGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCG
CTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGG
TAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCA
GGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATC
GACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCC
TCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGG
CGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGC
ACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGAC
ACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAG
AGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGC
CAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTT
TTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGT
CTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCT
AGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTT
ACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCC
CCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACC
CACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTG
CAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATA
GTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCA
GCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCG
GTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATT
CTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAAT
AGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTT
TAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCA
GTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAG
CAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCT
TCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTT
AGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTA
TTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACG
GTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGAC
AAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGA
TTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGA
AATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATA
GGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTG
GAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGG
CCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCC
TAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGC
GAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCT
TAATGCGCCGCTACAGGGCGCGTCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGG
GCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGG
TTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGTAATACGACTCACTATAGGGCGAATTAGTG
GATGAGAAGACCTGCAGAAAGAATTC (SEQ ID NO:27)

Figure 9

8 space
TCTAGAGAGAAAGGCTACGAAGATCCGGACATCAAAAGATCAAAACATGCGGAGATCGAC
ACATCCAAAAATGAACAGATAACAGGATCAGAGGGTCAAGAAGTGAAACCATCGAGGAGT
CGCAGACTACGAAGATGGACAAGTGAGCAACAGCTGCGATGGAGAACTAAAGGAGTCGGG
AGGTACGGCAGTGAAGAAATGTGCACATGAAAAGCTGCCCGGGTGCGCAACTGGAGGACT
AGAAGAATGGCACAGTGACCCGGTGAGAGCCTGAGGAGGTGCGAGAATAGGGCGATAGGG
AGGTAACCCGATGGCGGACTAAAGAGATACGGACATGGACGGATAGGAGAGTCGCAAGAT
GGAAGGATCACAAAGTGAGCCGGTCCAAGCATACGCGGGTAAACGGATCAGAACGTAAAG
CGGTGGAAAAATACCCGAGTAGGAAGATAGCAAAATACCCAGGTCGAGCGATAGCCCACT
CGGCGAGTCCAAAACTGAAAAAATGGAGAGATGAGACACTAGCGGCATCGGACAGTAGAG
CAGTGAACGCATCAGAGAGTAGAAAAGTGAAGAACTGCGACGGTGAAACCATAACAGAGT
CAGAGGGTACACCCATACGACCCTACCGCAATACAAGGCTAAGACGGTCAGAGGATGGAC
GGCTCGCGAACTCACGCCGTGCCAAAGTGAAGAAGTACCAAAATAGGAGCGTAACGGCGT
AGAGAGGTAGACGAGTGAGACGGTGGGCAGCTCCCAAGGTCCGAGCATGGACAGGTGGCA
AAATCCAGCGGTCAGAAACTACGAGACTGGCCGGATGAAGAAATAAGGCGCTAGAGAGAT
ACGGGAGTGACAAGATGGACCACTAGCGAAGTCGGAGGGTGAAGACGTGAACAACTAGAC
CAGTACGGCACTAGCAGAATAGACCGATGAAAACCTGAAAAGATAACAGAGTGAAGGCAT
CGGAGGGTCGAGAAGTAAGAACGTGGAGCCATACACGAATAAAGCAGTACAAGGGTGGAC
AAATGAGGAAATCGGGAGGTCCCACGGTGAAAACGTAGGGAGGTGAACCAGTGGGAGAGT
CAACCAATACGAAGATGCGAGGATCC (SEQ ID NO:28)

Figure 10A

10 space
CCATGGGAAAATGAGAGCCGCTCGACAAGGATAAAGAAAAGTCGCCCAACATCCGCCACG
CTAAGCGGCCCTAGGAGCCGGTAAACAAGCATGAAGAAACATGCGAGGCAATGACACAAA
CTCCAGGCGGGTGGAAAAGACTCGACGGAGATGCCACGCAATAGCGAAACCTGGAAGAGA
GTGGAAGCAAATAGGAAACAGTAGAGCAAAATGCCGCAGGATAGAAAACAGTGGCGCGCA
ATAAGCAGGAATAAAGCACAATAAAAAGAAGTAAACCGCACTAAAACAGCATCCAGACCG
CTGAAAAGAAATGGCGGCGAATGAAAAAAAGTCACAACCGGTGCAACAACCTGAAAAAGG
GTAGAGAGCCCTCCCGGGCAATAAAAGACGGTAAGAGGAAATCCCGAAAAGTGCAAAAAG
GTCAAAACGGGTGCAACAAAATGCCAAGGCGTGCAAGAAACTGAGGACGGATAAAAGACA
ATGCGGACGGGTACGACACGATAAGGAAAGATAAGGCGAAGTAAAAGCCGGTGCGGCAAA
ATACAACAGGCTGCCCAGAAGTACCGAGCGCTAGGCGGGAATGAGCGAGGGTAAAAGAGA
ATACACAACCCTGCAAGAAAGTCCGGGAAGCTAGGACCGGATACAGAGGAATAAGGAAGG
GTACACAGCAATACGAGCAGATGAACGAGCATACCAGAGAGATATCAGGCCTAAAAAACG
GTAACAAGAACTAGGAGGAGGTGCGAAGGCCTGGACCAAAATGGGACGAAGTAGGAGGGA
ATACACAACGATGGGACGAGATAGAGCCGGGTAGGCCAAAGTAGCACGGGATACCACAGA
ATACAGCGACGTACCGGAAAGTAGGAGAGACTGGAAACCCGTAAGGCAAGATAAGCGAGG
GCTAGCAGAACTAAAACAGACTGAACACGGATCAAGGCGAGTAGGCACGAGTACACGCCG
ATACGGACAAGTAGGCCGCAGTACGGAGAAGTAAACAGGGATGAGAGACAGTGAACAAGG
CTGAGACGAAATCACAGCAAATACAACAAAGTAGCGGCAAATAAACGGGAGTGGAACGCG
ATAGCAGGAAGTGACGGACAGCATGC (SEQ ID NO:29)

Figure 10B

Figure 10

12 space
GAATTCGAGTACCACCTGATGAGCAGCAAGGTAACGGCGAAGATGAAAAGAAAAATGGA
CGGCACGTACAAAAAAGGATGGCACCAGACATAGAAGAGGAGCTCAGCAAACGGCTGGA
AGCGAAGTCCCAGGAAGCCTACCAAGGACAGATATCAAAGAAATAAGAAACAAACTGAC
AAGCGCGTCGGAAGCGACGTACAAGCGGCAATGGCAGAGAGAGTAAAACAAACAGTGAA
CCAACAATAAACGCCCAGGTAGGGCGAAAGGTAAACAAAAGGATCGAAAGACGAGTAAC
CGGGAGATGGAAACAACGGTGCGAAAACCAGTGAAAAAGGGAGTCCAAAGCACCATGAA
ACCACAATACGACACACCCTGGGAGGACACGTGGAGGAGCCGGTACGAAGCAGGGTACA
GACAACATGCAAACAGAGGTAAAGGAACCGATAAAGAACGACATGGAAGGAAACGTAGC
ACCCGAGTAGAGGAACGGATAGCGAACGACGTGCAACGCCACCTAGAACGAGAAGTCGC
AGGGAAATAAACGGAGAGGTCCCACGGCCGATGAGCAGAAAAGTGGAAAAGGACCTGAA
AGAAAAATGAGGACCACAATAAAGGACGGAATAAAAAACACACTCCCAGAGGAAATACA
CGAGAAATAGGGAACAGGATGCAAAGAAGCATCAGAAAGGCAGTGCCAAAGGGAATCGC
CGCAAGCTAAGACAGGCCCTAGACCGCGAAGTAAACGAACAACTCGGAGGGCGAATCAC
CCCACCATGCAGAGCAAAATAAGAGACCAGATGCGCGGCAACCTGAGACCGAACCTAAA
GAAAGGATAGAGGGAGAGGTCCAAGAAACGATGGCGGGCACCATGAGAAAGGAAGTCAA
GTGCACCTAGCAACACAGATACCACGAGACATCGAAGCGAAGATACGGAAAAAGCTGGA
AGGGCAATGAACAGAAAGCTACGGAAAGAAGTCAAGGACACAATGAGGGAGCAAGTCGA
AGCGAGGTGGCCCGCCAAATAGGAGCAACAATGCGAGGCCAAGTGGGCACAACGATGAA
GACCAACTAACACAGCCACTTCTAGAGTGCACGGATCCCTCGAGGGTACCAACCATGGA
GCTAGCAAGCATGCAAACTGCAGAAGCTT (SEQ ID NO:30)

Figure 11A

14 space
CTCGAGGAAAGACAGAGGGTAGCAGGCCACAGATCAAGGCGGCGAAATGGCCCGGAGCA
GTAAGCACCACACGGTCACCAAGAGAGAATCAGAAGAGGCAAGTAAAAAAAAGCGCGTA
GGCACCCAAGGTGGACGCACAACAGTCACGGGAAAGAAGTCGAAGGCGGGCGCTCAGAA
AGCGAAGTACACAACCAGACGGTACCGGAACACGATGCGCCAACAAGAATCAGCGGAAA
ACATACCGGCGGACGGGTCGAAACACGAAAATGAAACCGCAGAAGTAAAGAGAAAAGGA
CGCAGGAAGAGGATCAAGACAGAAAAATCGACAAAGCAAAGTGGACCGAAGAAGATCGA
AAGAAGCACTCAGGACAGCTGAATCGCACAGACACGGTGACAAGACGAAAATGGACGAC
ACCGATAAGCGAAAAAAGATACCGGCGAGCGAGTCGAGAACACGAAGTCAAGGCGCAAG
ATAGAGCAAACGAAGTAAACAGGACAGAATAGGAAGGGCAAAGTGGGAGAGAAAGGCTC
GAACCGACGACTGCACCGAACAACGTGGGAGCCCACGGCTACAAAACGCCAGATCGAAC
CCAGGAGTCGAAGAACAAAAGTAACGGCGCAGAAATAGAGAGAGGAGGCTAAAGCGGAA
CAGTGAAGGAAAAGAAATAACGGGCCCAAGGTGAAGAAAAAGAAGTGGCAGGAACGGAG
ACAGCCCAGCAACTGGGCAAAAGCGGCTGACGAGAACAAGATAGAGAACGAGGAATGAC
AGACACAGATAGAAGACAAAGACTGCGGACAGGCAAGTAAGAGAAGGAAAATAAAGCGA
AGAAGTAGGCCGCCCGAAATGCAAAAGAAAAAGTGACACAGAGAAAATGACCGGCGGGA
GTAACACGCCCAGGATGACGGAGCACCAATAGAAAAAAACACGTAGGAGAGGGAACATC
GGACAAGACAATCCCAACCCGAAGATAAACACCGAGAGGTAGACAGCGCAGACTGGAGC
ACAGCCGTGGAAAGGGAAAGGTACGCAAACGGACCTAAAAAAAACAAAGTCAGGGAAAA
GGCTAACGGAACAAAGCTGGCCATGG (SEQ ID NO:31)

Figure 11B

Figure 11

16 space
GAATTCTCTAGAGTGCACGGATCCCTCGAGGGTACCAACCATGGAAGCTAGCAAGCATGC
GAGGAAAGCAAGGAAAATGAAGAACGAAGCGGGTCGAACAGGGAAGACATAAAAAACAGG
CACCATGACACCAACGCGGAATGACACAACACAAAGGTAGCGAGGCACGCACGTAAGAGG
CACAAGAAATGAAAGAAAAAGAGAATACGCAGAGAGGGCAGTAAGGCAGGGAAACGATGA
GCAGACGACGGGATGCTAGCAGCGAAAGCTGAGAAGAAGAGACCCTCAAAAAGCAAAAGA
ATAAGAAAGCCAAAGCATAAAAGAAAAGCGCAATCAAGGAAACGGGCGCTAGAAGACCGA
GACAGTGAGACGAGCACACGCTAAAAAAGAGAGACAATCCAGAAAGAAGACAGTGGACAA
GAGACCGGGTAGCACCAGAGGACAATGAAGGCGGGACGGCATAAACGCCCAAACGGGTAA
GACGGGCCAACGATAAAGGAAGACGAAGGTCAACGCAGAAGGGAATCCAGAAACACAGAC
ATAAAGAAAAAAAGAAATGGCAGCCGGAGAAGCTGAGACGGGACCGCAGTGCGAGAGAAG
AGACGTGCGGAAAAGAGAAGGTCGCGCAAGACGGGCGTGGACCGGACAAAGGATGGGAAA
ACGACCGGATGGAGAGCGCCCAGAGTCAGAAACGGAAGAGATGAAACCCAGAAAGACTCG
CCACAAGAAGGAGTCGGGAAACACAAGAATAAAGCAGCTGGGGAATAGCGGACGAACGGA
ATAGAGGCCCAAAGAACTGGGAGAGGACAGCCGTACAGGAGGAACGGGATAGCGAAGAAC
CAACGTGCGAAAAGACCGGCATAAAACAGAAAGGCAATAGCAGACAGCAAAAATAGAGAG
ACAACGACGTAGAAAGAGGGAAACGTCGCAAAAAACCGGGCTAGAGGAGGACCAACATAG
CGAGACACCAAGATGAAGCAACGGAGAAGTCACACAAAGAGACAGTCGAAAAGCAAAGAA
ATAGCCCAGCAAGCGAATACCCAACAAGAAGAGTCAACAGGAGAAAGAGTCCAAGGAGAA
GCAAGTACACAAGACAAAAGGTAGCAGACACAAGCAATGCACGGGCAGGGAAATGCGAAA
CAGAGAGGATAAGAAGATCTAAAATGCATGGTCTGTGTGATGTTGGTCTGTGTGATGTTG
GTCTGTGTGATGTTGGTCTGTGTGATGTTCTGCAGAAGCTT (SEQ ID NO:32)

Figure 12A

24 space
GGATCCGACGGGAACAAAGAGAAAGCACAAGTAAAAGGCAACGACGCGCACGAGGTCAGA
ACGCAAACAGGAAGAAGAATGCAACAAAGGAGCGCCGAACGGATACGCGGGCCAACACGA
AGAGGAATGAAAGGGAAACGAAACAAAAAAGTGAAGAGGCCAAAAGAGCACAGCATGGAG
GAAGCCACGGCCGAAAAACTACACCGGGAAGGACCAAACGCGATGGCAGACAGAGCGCAG
ACAGGGATAGAGCCACCCAGGAAGCACGCAGTCAGGGAGAAAAGGGCACAGACACTACGA
AAAAGGGCGACAAGAGGGCTGGGAGAGGGAACGGGCAGAAACATAAAACCCAAGAGCCGA
CCAGAAGTAAAAGGCGATATCAACACAGACGTAAAGCAAGCAGAACAAGAGCACATAAGC
AAGAAGGAGAGGAGACACCTACCAGAGAGAAAAACCCGAACCGTGCGGAAAGGCGGGCGA
GAAGGGATAAAAGAGCACGAACCCAAGAGGATGAAAGGACAAGAGAGAACGAAAGTCCAA
CACAGAAAGGCCAAAGGAATAAAAAGCAACAGCCAAAAAAAAGTAAGAAAGCAAGACAAA
GCGAGAATGAAAAGCGAAGGCACAAAAGAAATAGACAGAGGGAACAGCGGAAAGATCCCA
CGGGCGACAGCAACCGGACTCCACGGGAACACGAGGAGAAACGTGAAAACCGAGCGGCAA
CCCAGACTCAGCGGAACAGAAGACAAAAAAATACGGGCAGGGCGAGGAAAGAAAATAGGG
CCAACAAAAAGGAAGCAGATGAAAAACACAAAAACCCACCACATCACGCGGGCCGAGAGC
AGCAACGTGGCAGCACGAGGAGGAGGAGCCCTAGAACGCAAAGGCGAAAAGACGCTGCGA
GGTACCGGGCCAAAGAAAATAAAGGAACGCAAACAAAGGAGCATCAAAGGGAAGAAACAA
AAGAAAATAAAAAAGGAGAAGAAAGGGAAGATAAAAGGAAACGGAAAAAAGAAACTCGCG
ACAAAGACGGCAGCAAAGGTAGAGGAAAAGGAAGAGGACACACTGAAAGAGCACCAGGAG
AAAGACCTAGCAGAAGCACGCTCGAG (SEQ ID NO:33)

Figure 12B

Figure 12

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2009/053790

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68

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)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	GEISS GARY K ET AL: "Direct multiplexed measurement of gene expression with color-coded probe pairs" NATURE BIOTECHNOLOGY, NATURE PUBLISHING GROUP, NEW YORK, NY, US, vol. 26, no. 3, 1 March 2008 (2008-03-01), pages 317-325, XP002505107 ISSN: 1087-0156 the whole document	1-63
X	FORTINA P ET AL: "Digital mRNA Profiling" NATURE BIOTECHNOLOGY, NATURE PUBLISHING GROUP, NEW YORK, NY, US, vol. 25, no. 3, 1 March 2008 (2008-03-01), pages 293-294, XP008102233 ISSN: 1087-0156 the whole document	1-63

-/--

[X] Further documents are listed in the continuation of Box C.

[X] 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 document 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
20 November 2009	07/12/2009

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Costa Roldán, Nuria

Form PCT/ISA/210 (second sheet) (April 2005)

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1196

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2009/053790

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2007/139766 A (NANOSTRING TECHNOLOGIES INC [US]; HWANG JENQ-NENG [US]; MITTON JEFFREY) 6 December 2007 (2007-12-06) the whole document	1-63
X	WO 2007/076128 A (NANOSTRING TECHNOLOGIES INC [US]; INST SYSTEMS BIOLOGY [US]; GEISS GAR) 5 July 2007 (2007-07-05) the whole document	1-63
X	WO 2007/076132 A (NANOSTRING TECHNOLOGIES INC [US]; FERREE SEAN M [US]; DUNAWAY DWAYNE L) 5 July 2007 (2007-07-05) the whole document	1-63
X	US 2005/064435 A1 (SU XING [US] ET AL) 24 March 2005 (2005-03-24) the whole document	1-63
P,X	WO 2008/124847 A (NANOSTRING TECHNOLOGIES INC [US]; GEISS GARY K [US]; DAHL TIMOTHY [US]) 16 October 2008 (2008-10-16) the whole document	1-63
P,X	MALKOV VLADISLAV A ET AL: "Multiplexed measurements of gene signatures in different analytes using the Nanostring nCounter Assay System." BMC RESEARCH NOTES 2009, vol. 2, 2009, page 80, XP002556189 ISSN: 1756-0500 the whole document	1-63

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1197

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2009/053790

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.b of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application and necessary to the claimed invention, the international search was carried out on the basis of:

   a. type of material

   [X] a sequence listing

   [ ] table(s) related to the sequence listing

   b. format of material

   [X] on paper

   [X] in electronic form

   c. time of filing/furnishing

   [X] contained in the international application as filed

   [X] filed together with the international application in electronic form

   [ ] furnished subsequently to this Authority for the purpose of search

2. [ ] In addition, in the case that more than one version or copy of a sequence listing and/or table relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1198

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2009/053790

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2007139766 A	06-12-2007	AU 2007268027 A1	06-12-2007
		CA 2653095 A1	06-12-2007
		EP 2030011 A2	04-03-2009
WO 2007076128 A	05-07-2007	AU 2006330830 A1	05-07-2007
		CA 2640385 A1	05-07-2007
		EP 1963531 A2	03-09-2008
		JP 2009521230 T	04-06-2009
		WO 2007076129 A2	05-07-2007
WO 2007076132 A	05-07-2007	AU 2006330834 A1	05-07-2007
		CA 2635215 A1	05-07-2007
		EP 1963500 A2	03-09-2008
		JP 2009521688 T	04-06-2009
US 2005064435 A1	24-03-2005	CN 1856580 A	01-11-2006
		EP 1668164 A2	14-06-2006
		JP 2007506431 T	22-03-2007
		KR 20060094531 A	29-08-2006
		US 2006199216 A1	07-09-2006
		US 2007054288 A1	08-03-2007
		WO 2005030996 A2	07-04-2005
WO 2008124847 A	16-10-2008	NONE	

Form PCT/ISA/210 (patent family annex) (April 2005)

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 April 2011 (21.04.2011)

PCT

(10) International Publication Number
WO 2011/047087 A2

(51) International Patent Classification:
G01N 33/68 (2006.01) *G01N 33/58* (2006.01)
G01N 33/53 (2006.01) *C12Q 1/68* (2006.01)

(21) International Application Number:
PCT/US2010/052556

(22) International Filing Date:
13 October 2010 (13.10.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/251,192 13 October 2009 (13.10.2009) US
61/325,224 16 April 2010 (16.04.2010) US
61/326,787 22 April 2010 (22.04.2010) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:
US 61/325,224 (CIP)
Filed on 16 April 2010 (16.04.2010)
US 61/251,192 (CIP)
Filed on 13 October 2009 (13.10.2009)
US 61/326,787 (CIP)
Filed on 22 April 2010 (22.04.2010)

(71) Applicant *(for all designated States except US)*: **NANOSTRING TECHNOLOGIES, INC.** [US/US]; 530 Fairview Avenue North, Suite 2000, Seattle, WA 98109 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **DUNAWAY, Dwayne, L.** [US/US]; 2751 N.E. 94th Street, Seattle, WA 98115 (US).

(74) Agents: **ELRIFI, Ivor, R.** et al.; Mintz Levin Cohn Ferris Glovsky And Popeo, P.C., One Financial Center, Boston, MA 02111 (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, 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, 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, PE, PG, PH, PL, PT, RO, RS, RU, 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,

[Continued on next page]

(54) Title: PROTEIN DETECTION VIA NANOREPORTERS

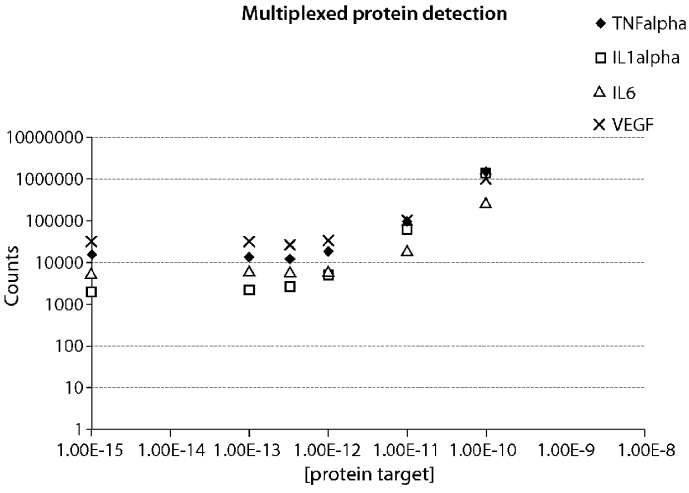


Fig. 10

(57) Abstract: The invention provides methods, compositions, kits and devices for the detection of proteins. In some embodiments, the invention allows for multiplexed protein detection.

WO 2011/047087 A2

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1200

GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

PROTEIN DETECTION VIA NANOREPORTERS

RELATED APPLICATIONS

[0001] This application claims the benefit of provisional applications USSN 61/251,192, filed October 13, 2009, USSN 61/325,224, filed April 16, 2010, and USSN 61/326,787, filed April 22, 2010, the contents of which are each herein incorporated by reference in their entireties.

FIELD OF THE INVENTION

[0002] The present invention relates generally to field of protein detection, quantification, identification, and multiplex analysis using the tools of molecular biology to generate unique nanoreporter constructs and the methods for using them.

BACKGROUND OF THE INVENTION

[0003] With the recent completion of analysis of the human genome, much attention is now shifting to the field of proteomics, where gene products (proteins), their variants, interacting partners and the dynamics of their regulation and processing are the emphasis of study. Such studies are essential in understanding, for example, the mechanisms behind genetic/and environmentally induced disorders or the influences of drug mediated therapies, as well as potentially becoming the underlying foundation for further clinical and diagnostic analyses. Critical to these studies is the ability to qualitatively determine specific variants of whole proteins (e.g., splice variants, point mutations, post-translationally modified versions, and environmentally/therapeutically-induced modifications) and the ability to view their quantitative modulation. Moreover, it is becoming increasingly important to perform these analyses from not just one, but multiple biological fluids/extracts. There are limited methods of multiplexed protein measurement technologies due to the additional challenges inherent in protein samples.

[0004] However, measurement of proteins in biological fluid is difficult due to their inherent properties. Accordingly, there is a pressing need for rapid, sensitive, reproducible, and accurate analytical approaches for the analysis of proteins and their variants.

[0005] In order to analyze proteins of interest from- and in- their native environment, assays capable of assessing proteins present in a variety of biological fluids and/or extracts, both qualitatively and quantitatively, are needed.

SUMMARY OF THE INVENTION

[0006] The invention provides method and compositions for analysis of proteins. In some embodiments, the invention provides methods and compositions for the detection and/or quantification of proteins in a sample. In some embodiments, the invention provides methods determining the concentration of at least one protein in a sample comprising the steps of: (a) providing: (i) at least one protein, (ii) a first protein probe specific for a first region of said at least one protein, where the first protein probe contains a capture region, (iii) a second protein probe specific for a second region of the at least one protein, where the second protein probe contains a nanoreporter comprising a plurality of different detectable labels, and (iv) a matrix having attached thereto a moiety which is capable of binding to the capture region in the first protein probe; (b) forming at least a complex comprising the at least one protein, the first protein probe, the second protein probe and the moiety, where the at least one protein is bound to the first and second protein probes, and where the moiety is bound to the capture probe in the first protein probe; and (c) individually detecting the complex or at least part of the complex by a method comprising individually counting the presence of one or more molecules of the nanoreporter where the presence of the one or more molecules is indicative of the concentration of the protein in the sample. In some embodiments, the individually detecting further comprises detecting a digital signal.

[0007] A moiety refers to and is also known as an entity. A moiety of the invention is operably linked to a matrix and binds with a capture region of a first protein probe. The moiety is operably linked to the matrix by a physical or chemical bond, including, but not limited to, a covalent bond, a non-covalent bond, an electron bond, a bent bond, an aromatic bond, a metallic bond, a hydrogen bond, an ionic bond, or van der Waals forces. The moiety binds with a capture region of a first protein probe through any of the physical or chemical bonds described herein, receptor-ligand interactions, hybridization events between two oligonucleotides, or interactions between an oligonucleotide and a polypeptide. For example, a capture region that contains biotin binds to a moiety containing streptavidin, forming a strong non-covalent bond, wherein a matrix having attached to the streptavidin, permits the matrix to bind to the capture region of the first protein probe (see, Figure 1). While all known receptor-ligand interactions are contemplated, those interactions with a dissociation constant (Kd) of between 0.1 fM and 1000 nM are preferred. Hybridization events occur between oligonucleotides having complementary sequences,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1203

however, perfect or complete complementarity is not required. The invention encompasses those hybridization events between oligonucleotides having 50%, 60%, 70%, 80%, 90%, 95%, 100%, and any percentage complementarity in between. Furthermore, the association of an aptamer with a first protein probe provides a non-limiting example of a preferred interaction between an oligonucleotide and a polypeptide.

[0008] In some embodiments, the invention provides methods for determining the concentration of a plurality of target proteins by forming a plurality of complexes, each complex comprising (i) at least one target protein (ii) a first protein probe specific for a first region of the at least one protein, where the first protein probe comprises a capture region (iii) a second protein probe specific for a second region of the at least one protein, where the second protein probe comprises a nanoreporter comprising a plurality of different detectable labels and (iv) a moiety attached to a matrix, where the moiety is capable of binding to the capture region in the first protein probe, where each second protein probe comprises a different nanoreporter region. In some embodiments, each nanoreporter in the plurality of complexes has a detectable signal that distinguishes it from other nanoreporters in the population. In some embodiments, the dissociation constant of the first and the second protein probes is about $1.00x10^{-10}$ to about 1.00x10-08. In some embodiments, the concentration of two or more target proteins is determined. In some embodiments, the concentration of 3, 4, 5, 10, 20, 30, 50, 100, 200, 300, 500, 600, 700, 800, 900, 1000 or more than 1000 different target proteins is determined. In some embodiments, the concentration of at least 972 different target proteins is determined.

[0009] In some embodiments, the matrix is selected from the group consisting of a bead and an array. In some embodiments, the matrix is a bead. In some embodiments where a plurality of target proteins is analyzed, the matrix is a bead and each moiety in each complex of the plurality of complex is attached to a different bead. In some embodiments, the matrix is a surface. In some embodiments where a plurality of target proteins is analyzed, the matrix is a surface and each moiety in each complex of the plurality of complex is attached to a different location of the surface.

[0010] In some embodiments, the first protein probe and the second protein probe are independently selected from the group consisting of antibody, peptide, aptamer and peptoid.

[0011] In some embodiments, the nanoreporter comprises a single-stranded nucleic acid backbone, the backbone comprising a plurality of label attachment regions covalently attached

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1204

together in a linear combination, where each label attachment region is hybridized to a complementary polynucleotide sequence having attached thereto the detectable label. In some embodiments, the nanoreporter is attached to the second probe through hybridization to a linker oligo. In some embodiments, the nanoreporter is hybridized to the linker oligo at a temperature of about 32 degrees Celsius (°C) to about 40°C. In some embodiments, the nanoreporter is hybridized to the linker oligo at a temperature of about 37°C. In some embodiments, the nanoreporter comprises a portion that is complementary to the linker oligo. In some embodiments, the complementary region is about 15 to about 20 bases.

[0012] In some embodiments, the invention provides methods for determining the concentration of at least one protein in a sample comprising the steps of: (a) providing: (i) at least one protein, (ii) a first protein probe specific for a first region of the at least one protein, where the first protein probe is attached to a first capture region or a first matrix, (iii) a second protein probe specific for a second region of the at least one protein, where the second protein probe comprises a signal oligo, and (iv) when the first probe is attached to a first capture region: a second matrix having attached thereto a moiety which is capable of binding to the capture region in the first protein probe; (b) forming at least a first complex comprising the at least one protein, the first protein probe, and the second protein probe, where the at least one protein is bound to the first and second protein probes, and where when the first probe is attached to a first capture region the capture probe is bound to the moiety in the second matrix; (c) releasing the signal oligo from the first complex; (d) forming a second complex comprising: (1) at least the signal oligo and (2) at least one oligo probe comprising a signal oligo-specific region and a region comprising a nanoreporter where the nanoreporter comprises a plurality of different detectable labels; and (e) individually detecting the second complex or at least part of the second complex by a method comprising individually counting the presence of one or more molecules of the nanoreporter, where the presence of the second one or more molecules is indicative of the concentration of the protein in the sample. In some embodiments, individually detecting further comprises detecting a digital signal.

[0013] In some embodiments, the first matrix is a bead or an array. Preferably, the first matrix is a bead. In other embodiments, the second matrix is a bead or an array.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1205

[0014] In some embodiments, the signal oligo is attached to a second capture region. In some embodiments, the releasing of the signal oligo further comprises capturing directly or indirectly the signal molecule into a third matrix.

[0015] In some embodiments, the nanoreporter further comprises a constant region, where the constant region comprises a plurality of repeat nucleotide sequences. In some embodiments, the constant region is bound to a second moiety in a third matrix, where the second moiety is capable of binding the constant region.

[0016] In some embodiments, the invention provides methods for determining the concentration of a plurality of target proteins by forming a plurality of complexes, each complex comprising (i) at least one target protein (ii) a first protein probe specific for a first region of the at least one protein, where the first protein probe is attached to a capture region or a first matrix (iii) a second protein probe specific for a second region of the at least one protein, where the second protein probe comprises a signal molecule, where when the first probe is attached to a first capture region the capture probe is bound to the moiety in the second matrix, and where each second protein probe in each the plurality of complexes comprises a different signal oligo. In some embodiments, the concentration of two or more target proteins is determined. In some embodiments, the concentration of 2, 3, 4, 5, 10, 20, 30, 50, 100, 200, 300, 500, 600, 700, 800, 900, 1000 or more than 1000 different target proteins is determined. In some embodiments, the concentration of at least 972 different target proteins is determined.

[0017] In some embodiments, the first matrix of the complex of the plurality of complexes is a bead and the bead comprises a plurality of identical first protein probes. The term identical is meant to describe a protein probe having the same sequence and either containing or attaching to the same capture region.

[0018] In some embodiments, the first protein probe and the second protein probe are independently selected from the group consisting of antibody, peptide, aptamer and peptoid.

[0019] In some embodiments, the nanoreporter comprises a single-stranded nucleic acid backbone, the backbone comprising a plurality of label attachment regions covalently attached together in a linear combination, where each label attachment region is hybridized to a complementary polynucleotide sequence having attached thereto the detectable label.

[0020] In some embodiments, the invention provides methods for determining the concentration of at least one protein in a sample comprising the steps of: (a) providing: (i) at least one protein,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1206

(ii) a first protein probe specific for a first region of the at least one protein, where the first protein probe is attached to a first oligo, and (iii) a second protein probe specific for a second region of the at least one protein, where the second protein probe is attached to a second oligo; (b) forming a first complex comprising the at least one protein, the first protein probe and the second protein probe, where the at least one protein is bound to the first and second protein probes; (c) ligating the first and the second oligo to form a signal oligo; (d) forming a second complex comprising: (1) the first signal oligo and (2) at least one oligo probe comprising a signal oligo-specific region and a region comprising a nanoreporter where the nanoreporter comprises a plurality of different detectable labels; and (e) individually detecting the second complex or at least part of the second complex by a method comprising individually counting the presence of one or more molecules of the nanoreporter, where the presence of the one or more molecules is indicative of the concentration of the protein in the sample. In some embodiments, the individually detecting further comprises detecting a digital signal.

[0021] In some embodiments, the signal oligo is released from the first complex. In some embodiments, the signal oligo comprises a capture region. In some embodiments, the releasing of the signal oligo further comprises capturing directly or indirectly the signal oligo into a matrix.

[0022] In some embodiments, the invention provides methods determining the concentration of a plurality of target proteins by forming a plurality of complexes, each complex comprising (i) at least one target protein, (ii) a first protein probe specific for a first region of the at least one protein, where the first protein probe is attached to a first oligo (iii) a second protein probe specific for a second region of the at least one protein, where the second protein probe is attached to a second oligo, where the ligation of the first oligo and the second oligo form a signal oligo, and where each complex in the plurality of complexes comprises a different signal oligo. In some embodiments, the concentration of two or more target proteins is determined. In some embodiments, the concentration of 2, 3, 4, 5, 10, 20, 30, 50, 100, 200, 300, 500, 600, 700, 800, 900, 1000 or more than 1000 different target proteins is determined. In some embodiments, the concentration of at least 972 different target proteins is determined.

[0023] In some embodiments, the first protein probe and the second protein probe are independently selected from the group consisting of antibody, peptide, aptamer and peptoid.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1207

[0024] In some embodiments, the nanoreporter comprises a single-stranded nucleic acid backbone, the backbone comprising a plurality of label attachment regions covalently attached together in a linear combination, where each label attachment region is hybridized to a complementary polynucleotide sequence having attached thereto the detectable label.

[0025] In some embodiments, the invention provides a population of uniquely labeled protein probes, where each probe comprises: i) a target-specific region; and ii) a region comprising a nanoreporter comprising a plurality of different detectable molecules, where the nanoreporter in each protein probe has a detectable signal that distinguishes it from other nanoreporters in the population. In some embodiments, the target-specific region is selected from the group consisting of antibody, peptide, aptamer and peptoid.

[0026] In some embodiments, the nanoreporter comprises a single-stranded nucleic acid backbone, the backbone comprising a plurality of label attachment regions covalently attached together in a linear combination, where each label attachment region is hybridized to a complementary polynucleotide sequence having attached thereto the detectable label. In some embodiments, the dissociation constant of the target specific region is about $1.00x10^{-10}$ to about $1.00x10^{-8}$.

[0027] In some embodiments, the nanoreporter is attached to the protein probe through hybridization to a linker oligo. In some embodiments, the nanoreporter is hybridized to the linker oligo at a temperature of about 32°C to about 40°C. In some embodiments, the nanoreporter is hybridized to the linker oligo at a temperature of about 37°C. In some embodiments, the nanoreporter comprises a portion that is complementary to the linker oligo. In some embodiments, the complementary region is about 15 to about 20 bases.

INCORPORATION BY REFERENCE

[0028] All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0029] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1208

by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0030] Figure 1 is a schematic diagram depicting one embodiment of the invention in which two antibodies specific for a target protein bind to the target protein in solution. The first antibody is attached to an affinity tag such as biotin (depicted as a circle labeled "B"), while the second antibody is attached to a partially double stranded nucleic acid probe. The binding of the first and second antibodies to the target protein forms a complex that is isolated from the solution via the affinity tag of the first antibody. One of the strands of the partially double stranded nucleic acid probe can be eluted to generate a signal oligo, which can then be analyzed by any the methods described herein.

[0031] Figure 2 is a graph depicting the results of a detection assay using the IL-2 target protein at different concentrations. Specifically, the detection of IL-2 within solutions containing no blocker, milk at 0.03%, bovine serum albumin (BSA) at 0.1%, or salmon sperm (SS) at 98 ng/ml, was measured as the total counts detected as a function of increasing IL-2 target protein molar concentration ([IL2 target, M]).

[0032] Figure 3 is a graph depicting the efficiency of IL-2 detection in the assay used in Figure 2. Total counts detected were normalized to 600 molecules per field of view (FOV) and expressed as a function of increasing concentration of IL2 target protein molecules ([IL2 target, molecules]). The efficiency of detection is the slope of the line depicted in this graph.

[0033] Figure 4 is a schematic diagram depicting two alternate embodiments of the invention for solution tripartite binding. According to this method, two antibodies specific for a target protein bind to that target protein in solution. The first antibody is attached to an affinity tag such as biotin and contains a constant region, which, for example, contains F repeats. The second antibody is attached to a nanoreporter probe and a second constant region, which, for example, contains, G repeats. The binding of the first and second antibodies to the target protein forms a complex that can be isolated from the solution via the affinity tag of the first antibody. "Normal" elution of the complex is accomplished by melting off the G and F bead. "Alternative" elution of the complex is accomplished via digestion. The label monomers of the nanoreporter (depicted as circles) emit individual signals of qualitatively different wavelengths that are spatially-distinguishable and are, from left to right positions, red (R), yellow (Y), green (G), blue (B), red (R), and violet (V).

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1209

[0034] Figure 5 is a graph depicting the dissociation constant (Kd) and probe concentration for one of the embodiments of the invention, expressed as the fraction of target bound versus the Kd of the nanoreporter probe and the protein probe.

[0035] Figure 6 is a schematic diagram depicting an embodiment of the invention in which a capture antibody specific for a target protein binds to the target protein in solution to form a complex. The complex can then be isolated from the solution. The complex is then contacted with a second antibody, where the second antibody is attached to a partially double stranded nucleic acid probe. One of the strands of the partially double stranded nucleic acid probe can be eluted to generate a signal oligo that may be analyzed by any the methods described herein.

[0036] Figures 7A and 7B are schematic diagrams depicting an embodiment of the invention in which two antibodies specific for a target protein bind to the target protein in solution. The first antibody is a capture antibody, while the second antibody is attached to a partially double stranded nucleic acid probe, where one of the strands in the probe is attached to an affinity tag such as biotin. The binding of the first and second antibodies to the target protein formed a complex that can be isolated from the solution via the capture antibody. One of the strands of the partially double stranded nucleic acid probe can be eluted to generate a signal oligo containing the affinity tag. The signal oligo can then be hybridized to a nanoreporter to form a nanoreporter-signal oligo complex that can be isolated and/or immobilized into a solid surface. The nanoreporter-signal oligo complex can be analyzed by any the methods described herein. The label monomers of the nanoreporter (depicted as circles) emit individual signals of qualitatively different wavelengths that are spatially-distinguishable and are, from left to right positions, red (R), yellow (Y), green (G), blue (B), red (R), and violet (V).

[0037] Figure 8A is a schematic diagram depicting certain embodiments of the invention using proximity ligation. A first and a second oligo are attached to a first and a second antibody, respectively, both antibodies being specific for a target protein. The first and second antibodies bind to the target protein, bringing the first and second oligo to close proximity. A bridging oligo and a ligase are added to the solution to connect the first and second oligo to generate a signal oligo. The signal oligo can then be analyzed by any the methods described herein.

[0038] Figures 8B-D are schematic diagrams depicting methods by which the signal oligo shown in Figure 8A may be released and purified.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1210

[0039] Figure 9 is a schematic diagram depicting certain embodiments of the invention using proximity ligation. The label monomers of the nanoreporter (depicted as circles) emit individual signals of qualitatively different wavelengths that are spatially-distinguishable and are, from left to right positions, red (R), yellow (Y), green (G), blue (B), red (R), and violet (V).

[0040] Figure 10 is a graph depicting the results of multiplexed protein detection of TNFalpha, IL1alpha, IL6, and VEGF, measured as total counts detected as a function of increasing protein target concentration ([protein target]). In this example, a sandwich detection assay was used in solution. A 4-plex measurement is shown.

[0041] Figure 11 is a graph depicting the data analyzed in Figure 10, plotted by lane instead of by concentration. Specifically, this figure demonstrates that two target proteins were titrated in whereas two other protein targets were titrated out.

[0042] Figure 12 is a graph depicting the results of a limit of detection (LOD) experiment using two protein targets, IL1alpha and IL6. Total counts detected were plotted as a function of increasing molar concentration of the target protein ([target] molar). The experiment demonstrated that the limits of detection were 26 and 38 picograms per milliliter (pg/ml), corresponding to $1.4x10^{-12}$M and $1.9x10^{-12}$M for IL1alpha (IL1α) and IL6, respectively. The limit of detection was two standard deviations above background detection levels. Six negative controls were performed, resulting in average counts of plus or minus one standard deviation, *i.e.* 3196 ± 265 and 6703 ± 585, respectively.

[0043] Figure 13 is a graph depicting the total counts retained by various components of the antibody reporter complex (PROX01, PROX03, PROX04, PROX05, and PROX06) following purification and a rinsing step using either water or SSPE buffer of various fold concentrations (0.01X, 0.03X, or 0.1X). At 0.03X SSPE the oligo representing a ligated product, PROX05, was retained.

[0044] Figure 14 is a graph showing the counts per field of view (FOV) for each antibody probe on a reporter that bound, stretched, and immobilized (S17, S8, S22, S14, S23, S6, S13, S7, S18, S9, S10, S11, S12, S15, S16, S19, S20, and S21). Counts are shown only of the reporter with the antibody probe that is bound to the surface.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1211

DETAILED DESCRIPTION OF THE INVENTION

[0045] Reference will now be made in detail to particularly preferred embodiments of the invention. Examples of the preferred embodiments are illustrated in the following Examples section.

[0046] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which this invention belongs. All patents and publications referred to herein are incorporated by reference in their entirety.

[0047] The present invention provides compositions and methods for detection and quantification of individual target molecules in biomolecular samples. In particular, the invention provides protein probes that are capable of binding individual target molecules. The invention also provides the use of nanoreporters. Through nanoreporters' label codes, the binding of the protein probes to target molecules results in the identification of the target molecules. Methods of making and using such protein probes and/or nanoreporters are also provided. The methods and compositions described herein can be used in a wide variety of applications such as diagnostic, prognostic, quality control and screening applications.

[0048] Certain aspects of the invention relate to the detection of multiple target molecules. Multiplexing is the measurement of more than one target molecule within a sample without having to split the sample. The methods described herein provide potential benefits in the areas of multiplexing, quantification, and sensitivity. For example, in some embodiments the target molecule is a protein. Measurement of protein concentrations is challenging. Proteins are sticky and tend to aggregate. In addition proteins are unstable, and tend to unfold easier than RNA or DNA. Extremes in pH, temperatures, solute concentration, and the presence of denaturants are conditions that can interrupt protein stability and complicate measurement. In some embodiments, the invention provides methods and compositions for multiplexed protein measurements that are sensitive and reliable.

[0049] Multiplexing within a fluid sample is a key advantage of this approach. Multiplexing within one sample saves significant labor, reduces sample quantity requirements proportional to the number of measurements, and improves accuracy by elimination of errors compounded by separate sample handling and measurement steps. In some embodiments, the methods described herein allow for the pooling of different samples together during processing to be analyzed at

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1212

once. This offers throughput advantages and can accelerate the analysis of different samples, e.g., up to eight times.

[0050] In some embodiments, the invention provides protein probes for the analysis of target molecules. In some embodiments, the invention provides a protein probe population for use in a multiplexed assay. Each protein probe in the population is specific for a target molecule. The binding of the target molecules to the proteins probes is then detected using nanoreporters. Each nanoreporter comprises a unique label code that can be associated to a specific target molecule.

[0051] In some embodiments, the nanoreporters are attached, directly or indirectly, to the protein probes. A unique nanoreporter's label code is then assigned to a specific protein probe such that each nanoreporter's label code can be associated to the target molecule bound to the protein probe.

[0052] In other embodiments, the protein probes are attached, directly or indirectly, to a signal oligo. Each protein probe is attached to a unique signal oligo. The nanoreporters used for the analysis of the signal oligo comprise a portion that is complementary to the signal oligo. A unique nanoreporter's label code is assigned to a specific signal oligo sequence such that each nanoreporter's label code can be associated to the target molecule via the signal oligo sequence.

[0053] In other aspect of the invention, the invention provides methods for detecting target molecules by measuring signals digitally. Current technologies use analogue fluorescent signals to quantify the presence of target molecules. Quantification using fluorescence can be error prone for a variety of reasons. For example, fluorophores can photobleach. There can be changes in the spectra in the presence of proteins or due to local environment, e.g., pH, salt. In addition, the light sources can vary in intensity over time. For example, arc lamps, a commonly used light source, demonstrate a phenomenon called arc wander that can cause significantly different illumination levels over time. In embodiments of the invention, the target molecules are detected digitally. While fluorescence might be used to read the nanoreporter's label code, the signals are high and the spot is either present of not, thus the digital detection. The digital detection of target molecules leads to more accurate quantification.

Protein Probes

[0054] Protein probes are molecules or assemblies that are designed to bind with at least one target protein, at least one target protein surrogate, or both; and can, under appropriate

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1213

conditions, form a molecular complex comprising the protein probe and the target protein. The terms "protein", "polypeptide", "peptide", and "amino acid sequence" are used interchangeably herein to refer to polymers of amino acids of any length. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non amino acids or synthetic amino acids. The terms also encompass an amino acid polymer that has been modified, for example, by disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation, such as conjugation with a labeling component. As used herein the term "amino acid" refers to either natural and/or unnatural or synthetic amino acids, including but not limited to glycine and both the D or L optical isomers, and amino acid analogs and peptidomimetics.

[0055] The methods of the invention also encompassed protein probes designed to bind targets other than proteins. Examples of target other than proteins include, but are not limited to, nucleic acids, lipids, carbohydrates, ions, small molecules, organic monomers, and drugs. For convenience only, most of the embodiments described herein are explained in the context of protein probes that bind to a target protein. However, these embodiments also can be applied to other target molecules.

[0056] Protein probes typically are part of at least one probe set, comprising at least one first probe and at least one second probe. In certain embodiments, however, at least one probe set can comprise only first probes or second probes, but not both first probes and second probes. Probes comprise at least one reaction portion that allow them to bind to or interact with at least one target protein, at least one part of at least one target protein, at least one target protein surrogate, at least part of a target protein surrogate, or combinations thereof; typically in a sequence-specific, a confirmation-specific manner, or both; for example but not limited to antigen-antibody binding, aptamer-target binding, and the like.

[0057] In certain embodiments, the protein probes comprise an identity portion or at least part of an identity portion, for example, a signal oligo, a nanoreporter and/or linker oligo. In certain embodiments, the protein probes comprise a capture region. In some embodiments, the capture region is used for the isolation of the protein probe and/or immobilization of the protein probe into a surface. The capture region can be an affinity tag as described below, a bead, a slide or an array.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1214

[0058] In some embodiments, the protein probe is an antibody. As used herein, the terms antibody and antibodies are used in a broad sense, to include not only intact antibody molecules, for example but not limited to immunoglobulin A, immunoglobulin G and immunoglobulin M, but also any immunoreactive component(s) of an antibody molecule that immunospecifically bind to at least one epitope. Such immunoreactive components include but are not limited to, Fab fragments, Fab' fragments, $F(ab')_2$ fragments, single chain antibody fragments (scFv), miniantibodies, diabodies, crosslinked antibody fragments, Affibody™, cyclotides, molecules, and the like. Immunoreactive products derived using antibody engineering or protein engineering techniques are also expressly within the meaning of the term antibodies. Detailed descriptions of antibody and/or protein engineering, including relevant protocols, can be found in, among other places, J. Maynard and G. Georgiou, Ann. Rev. Biomed. Eng. 2:339 76 (2000); Antibody Engineering, R. Kontermann and S. Dubel, eds., Springer Lab Manual, Springer Verlag (2001); U.S. Pat. No. 5,831,012; and S. Paul, Antibody Engineering Protocols, Humana Press (1995).

[0059] The skilled artisan will appreciate that antibody can be obtained from a variety of sources, including but not limited to polyclonal antibody, monoclonal antibody, monospecific antibody, recombinantly expressed antibody, humanized antibody, plantibodies, and the like; and can be obtained from a variety of animal species, including rabbit, mouse, goat, rat, human, horse, bovine, guinea pig, chicken, sheep, donkey, human, and the like. A wide variety of antibody is commercially available and custom-made antibody can be obtained from a number of contract labs. Detailed descriptions of antibodies, including relevant protocols, can be found in, among other places, Current Protocols in Immunology, Coligan et al., eds., John Wiley & Sons (1999, including updates through August 2003); The Electronic Notebook; Basic Methods in Antibody Production and Characterization, G. Howard and D. Bethel, eds., CRC Press (2000); J. Goding, Monoclonal Antibodies: Principles and Practice, 3d Ed., Academic Press (1996); E. Harlow and D. Lane, Using Antibodies, Cold Spring Harbor Lab Press (1999); P. Shepherd and C. Dean, Monoclonal Antibodies: A Practical Approach, Oxford University Press (2000); A. Johnstone and M. Turner, Immunochemistry 1 and 2, Oxford University Press (1997); C. Borrebaeck, Antibody Engineering, 2d ed., Oxford university Press (1995); A. Johnstone and R. Thorpe, Immunochemistry in Practice, Blackwell Science, Ltd. (1996); H. Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); and S. Hockfield et

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1215

al., Selected Methods for Antibody and Nucleic Acid Probes, Cold Spring Harbor Lab Press (1993). Additionally, a vast number of commercially available antibodies, including labeled or unlabeled; polyclonal, monoclonal, and monospecific antibodies, as well as immunoreactive components thereof; custom antibody suppliers, and the like can be found on the World Wide Web at, among other places, the Antibody Search page at biocompare.com, the Antibody Resource Page at antibodyresource.com, and the Antibody Explorer page at sigmaaldrich.com.

[0060] In some embodiments, the antibodies described herein are attached to a nucleic acid, e.g., signal oligo, linker oligo and/or nanoreporter. Methods to attach nucleic acids to antibodies are known in the art. Any suitable method to attach nucleic acids to antibodies is encompassed in the methods of the invention. The antibodies described herein can be attached to a nucleic acid by the methods described in Gullberg et al., PNAS 101 (22): pages 228420–8424 (2004); and Boozer et al, Analytical Chemistry, 76(23): pages 6967-6972 (2004), both incorporated herein by reference. The antibodies described herein can be attached to a nucleic acid by random amine attachment. In some embodiments, the antibodies described herein can be attached to a nucleic acid by random amine attachment using a 10 to 1 ratio of nucleic acid to antibody. The antibodies described herein can be attached to a nucleic acid by the methods described in Kozlov et al., Biopolymers 5: 73 (5): pages 621–630 (2004) incorporated herein by reference. The antibodies described herein can be attached to a nucleic acid by hydrazine chemistry. The antibodies described herein can be attached to a nucleic acid using tadpoles as described in Nolan, Nature Methods 2, 11 - 12 (2005), incorporated herein by reference. The antibodies described herein can be attached to a nucleic acid by any suitable methods known in the art to generate engineered antibodies including the ones described herein.

[0061] In some embodiments, the protein probe is an aptamer. Aptamers include nucleic acid aptamers (i.e., single-stranded DNA molecules or single-stranded RNA molecules) and peptide aptamers. Aptamers bind target molecules in a highly specific, conformation-dependent manner, typically with very high affinity, although aptamers with lower binding affinity can be selected if desired. Aptamers have been shown to distinguish between targets based on very small structural differences such as the presence or absence of a methyl or hydroxyl group and certain aptamers can distinguish between D- and L-enantiomers. Aptamers have been obtained that bind small molecular targets, including drugs, metal ions, and organic dyes, peptides, biotin, and proteins, including but not limited to streptavidin, VEGF, and viral proteins. Aptamers have been shown

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1216

to retain functional activity after biotinylation, fluorescein labeling, and when attached to glass surfaces and microspheres.

[0062] Nucleic acid aptamers, including speigelmers, are identified by an in vitro selection process known as systematic evolution of ligands by exponential amplification (SELEX). In the SELEX process very large combinatorial libraries of oligonucleotides, for example 10^{14} to 10^{15} individual sequences, often as large as 60 -100 nucleotides long, are routinely screened by an iterative process of in vitro selection and amplification. Most targets are affinity enriched within 8 - 15 cycles and the process has been automated allowing for faster aptamer isolation. Peptide aptamers are typically identified by several different protein engineering techniques known in the art, including but not limited to, phage display, ribosome display, mRNA display, selectively infected phage technology (SIP), and the like. The skilled artisan will understand that nucleic acid aptamers and peptide aptamers can be obtained following conventional procedures and without undue experimentation. Detailed descriptions of aptamers, including relevant protocols, can be found in, among other places, L. Gold, J. Biol. Chem., 270(23):13581 84 (1995); S. Jayasena, Clin. Chem., 45:1628-50 (1999); V. Sieber et al., Nat Biotechnol. 16 (10):955-60 (1998); D. Wilson and J. Szostak, Ann. Rev. Biochem. 68:611-47 (1999); L. Jermutus et al., Eur. Biophys. J., 31:179-84 (2002); S S. Spada et al., Biol. Chem., 378:445-56 (1997); B. Wlotzka et al., Proc. Natl. Acad. Sci., 99:8898-8902 (2002).

[0063] In some embodiments the aptamer will be ligated or hybridized to a signal oligo, a linker oligo and/or a nanoreporter. In some embodiments, the ligation of the aptamer to a nanoreporter is done before annealing segments with labels to the nanoreporters. The hybridization or ligation of aptamers can be done by any suitable method known in art. For example ligation can be performed enzymatically by at least one DNA ligase or at least one RNA ligase, for example but not limited to, T4 DNA ligase, T4 RNA ligase, Thermus thermophilus (Tth) ligase, Thermus aquaticus (Taq) DNA ligase, or Pyrococcus furiosus (Pfu) ligase. Ligation can also be performed by chemical ligation can, using activating and reducing agents such as carbodiimide, cyanogen bromide (BrCN), imidazole, 1-methylimidazole/carbodiimide/cystamine, N-cyanoimidazole, dithiothreitol (DTT) and ultraviolet light.

[0064] In some embodiments, the protein probe is a peptoid. Peptoids are short sequences of N-substituted glycines synthetic peptides that bind proteins. In some embodiments, small size peptoids improve diffusion and kinetics of the methods described herein. Any suitable method

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1217

known in the art to generate peptoids is encompassed in the methods described herein. See Simon et al., PNAS 15; 89(20): 9367–9371 (1992), incorporated herein by reference.

Target Proteins

[0065] Target proteins are the protein detected or measured by binding of a protein probe whose target-specific region(s) recognize thereto. However, the invention encompasses detection of other targets beyond proteins such as nucleic acid, a lipid, a carbohydrate, a small molecule, an organic monomer, or a drug. Nucleic acids that can be analyzed by the methods herein include: double-stranded DNA, single-stranded DNA, single-stranded DNA hairpins, DNA/RNA hybrids, RNA (e.g. mRNA or miRNA) and RNA hairpins. For convenience only, the methods described herein are explained mostly in the context of analyzing proteins. However, the embodiments described herein also can be used to detect non-protein targets.

[0066] A target protein can be part of a biomolecular sample that contains other components or can be the sole or major component of the sample. A target protein can be a component of a whole cell or tissue, a cell or tissue extract, a fractionated lysate thereof or a substantially purified molecule. The target protein can be attached in solution or solid-phase, including, for example, to a solid surface such as a chip, microarray or bead. Also the target molecule can have either a known or unknown structure or sequence.

[0067] The compositions, methods, and kits disclosed herein can also be used in a wide variety of applications to determine the presence of target proteins in a sample. For example but without limitation, the compositions, methods, and kits are useful for, pharmacokinetic studies, including but not limited to, drug metabolism, ADME profiling, and toxicity studies; target validation for drug discovery; protein expression profiling; proteome analyses; metabolomic studies; post-translation modification studies, including but not limited to glycosylation, phosphorylation, acetylation, and amino acid modification, such as modification of glutamate to form gamma-carboxy glutamate and hydroxylation of proline to form hydroxylation; analyses of specific serum or mucosal antibody levels; evaluation of non-nucleic acid diagnostic indicators; foreign antigen detection; and the like.

[0068] In certain embodiment, at least one first protein probe, at least one second protein probe, or both the first protein probe and the second protein probe of at least one probe set comprise at least one antibody, aptamer or peptoid that reacts specifically with at least one target protein or at least one target protein surrogate. In certain embodiments, at least one first protein probe, at least

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1218

one second protein probe, or both the first protein probe and the second protein probe of at least one probe set comprise binding proteins that specifically interact with at least one target protein or at least one target protein surrogate.

[0069] The skilled artisan understands that with antibody probes, the reactive portion typically comprises the antigen binding site and related residues of the antibody molecule; and the target sequences comprise that portion of the analyte that includes the epitope, whether such sequences are linear, conformational, or combinations thereof. The skilled artisan will appreciate that the molecular complexes and the at least part of the molecular complexes described herein can be individually detected while tethered or attached to a substrate or while in solution, depending on, among other things, the nature of the specific molecular complex or cleavable component and the SMD technique and detection apparatus employed.

[0070] Protein isolation techniques are also well known in the art and kits employing at least some of these techniques are commercially available. Protein isolation techniques typically employ one or more of the following: maceration and cell lysis, including physical, chemical and enzymatic methods; centrifugation; separations by molecular weight, such as size exclusion chromatography and preparative electrophoresis; selective precipitation, for example, salting-in and salting-out procedures; various chromatographic methods; and the like. Detailed descriptions of and relevant protocols for protein purification techniques can be found in, among other places, Marchak et al., Strategies for Protein Purification and Characterization: A Laboratory Course Manual, Cold Spring Harbor Press (1996); Essentials from Cells: A Laboratory Manual, D. Spector and R. Goldman, eds., Cold Spring Harbor Press (2003); R. Simpson, Proteins and Proteomics: A Laboratory Manual, Cold Spring Harbor Press (2003); and D. Liebler, Introduction to Proteomics, Humana Press (2002). Commercially available kits can also be used, for example but not limited to, ProteoExtract.TM. Partial Proteome Extraction Kits (P-PEK) and ProteoExtract.TM. Complete Proteome Extraction Kits (C-PEK), available from CALBIOCHEM.RTM., La Jolla, Calif. The skilled artisan will appreciate that non-nucleic acid analytes for use with the inventive compositions, methods, and kits can be readily obtained without undue experimentation using such purification techniques and commercial kits.

Methods

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1219

[0071] The present invention provides methods for detection and quantification of individual target proteins in biomolecular samples. In particular, the invention provides protein probes that are capable of binding individual target proteins. The invention also provides the use of nanoreporters. Through nanoreporters' label codes, the binding of the protein probes to target proteins results in the identification of the target proteins. Methods of making and using such protein probes and/or nanoreporters are also provided.

[0072] In some embodiments, the invention provides methods for detection and/or quantification of a target protein by binding a protein probe to a target protein. A protein probe comprises at least one reaction portion that allow the probe to bind to or interact with the target protein or a target protein surrogate or combinations thereof; typically in a sequence-specific, a confirmation-specific manner, or both; for example but not limited to antigen-antibody binding, aptamer-target binding, and the like.

[0073] Protein probes typically are part of at least one probe set, comprising at least one first probe and at least one second probe. Thus, in some embodiments the invention provides methods for detection and/or quantification of a target protein by binding a protein probe set to a target protein, where the protein probe set comprises a first protein probe and a second protein probe. The first protein probe and the second protein probe comprise at least one reaction portion that allow the probes to bind to or interact with different regions of the target protein or a target protein surrogate or combinations thereof, e.g., in a sequence-specific manner, a confirmation-specific manner, or both.

[0074] In some embodiments, the methods described herein further comprise protein probes containing an identity portion or at least part of an identity portion, for example, a signal oligo, a nanoreporter and/or linker oligo. The identity portion allows for the identification of the presence or absence of the protein probe or probes bound to the target protein in the detection step of the methods described herein. Thus, in some embodiments the invention provides methods for detection and/or quantification of a target protein by binding the protein probe or protein probe set to a target protein, wherein the protein probe or at least one of the protein probes in the probe set contains an identity portion (e.g., a signal oligo, a nanoreporter and/or linker oligo).

[0075] In some embodiments, the identity portion is a signal oligo. A signal oligo comprises a polynucleotide sequence. Each protein probe or protein probe set will have a specific and/or

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1220

unique signal oligo in an assay, such that the signal oligo can be associated with the target protein. In certain embodiments, the signal oligo comprises about 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, 50, 60, 70 or more nucleotide bases In one embodiment, the signal oligo comprises between 40 to 120 bases, or between 80 and 100 bases. In some embodiments, the signal oligo is bioatinylated and used with a capture probe and a nanoreporter as described below. The signal oligo can be attached directly or indirectly to the protein probe. Methods for attaching nucleic acid to proteins probes are known in art including those described herein. Signal oligos can be a designed synthetic nucleic acid sequences or a natural sequence derived from a natural source such as sequence from viral genome, bacteriophages, or animal genomes.

[0076] In some embodiments, the signal oligo is attached indirectly to a protein probe through hybridization with a linker oligo attached to the protein probe. A linker oligo comprises a polynucleotide sequence. In the embodiments in which a linker oligo is used, each linker oligos will be specific and/or unique for a protein probe or protein probe set in an assay such that the complementary signal oligo can be associated to the target protein. The signal oligo comprises a portion that is complementary to the linker oligo attached to the protein probe. In some embodiments, the complementary portion of the signal oligo is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, 50, 60, 70 or more nucleotide bases. In some embodiments, the complementary portion of the signal oligo is 10-25 bases. In some embodiments, the complementary portion of the signal oligo is in the range of 15-20 bases. In some embodiments, the complementary portion of the signal oligo is 40 bases. In some embodiments, the complementary portion of the signal oligo is 30 bases. In some embodiments, the complementary portion of the signal oligo is 20 bases. The linker oligo can be a designed synthetic nucleic acid sequences or a natural sequence derived from a natural source such as a sequence from viral genome, bacteriophages, or animal genomes.

[0077] Figure 1 shows a schematic representation of one of the embodiments of the invention in which a signal oligo is used for the detection of the target protein. The embodiment depicted in Figure 1 is set up to separate the binding of the target protein from the hybridization of the nanoreporters. Figure 1 in step 1) shows a first protein probe comprising a signal oligo attached to the probe via hybridization with a linker oligo; and a second protein attached to an affinity tag. In the embodiment depicted in Figure 1 the protein probes are antibodies and the affinity tag is

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1221

biotin. However, the embodiment depicted in this figure can utilize any of the protein probes and affinity tags described herein. Both the first and second protein probes comprise a target specific region capable of binding one or more portions of a target. In step 2) and 3), the target protein is mixed with the first and second protein probes. In step 4), the complex of target protein and protein probes is purified. In the example depicted in Figure 1 the complex of target protein and protein probes is purified using streptavidin-coupled magnetic beads, such as Dynabeads® (Invitrogen). However, in this or any other embodiment described herein, the complex of target protein and protein probe (s) can be purified by any suitable method known in the art such as chromatography, including but not limited to HPLC, FPLC, size exclusion (gel filtration) chromatography, affinity chromatography, ion exchange chromatography, hydrophobic interaction chromatography, immunoaffinity chromatography, and reverse phase chromatography; ligand-receptor binding, such as biotin-avidin, maltose-maltose binding protein (MBP), calcium-calcium binding peptide; aptamer-target binding; zip code hybridization; and the like.

[0078] In step 5) of Figure 1, the signal oligo is eluted from the complex of target protein and protein probes and analyzed using nanoreporters as described below. Methods for eluting the signal oligos are know in the art including the ones depicted in Figure 1 and described herein. In some embodiments, the methods depicted in Figure 1 are used to detect and/or quantify a plurality of target proteins. Each target protein will be detected by a probe set comprising a first probe and a second probe as described in Figure 1. Each probe set will have a specific and/or unique signal oligo that can then be associated to the target protein of each probe set.

[0079] In some embodiments, the protein probes comprise a capture region. In some embodiments, the capture region is used for the isolation of the protein probe and/or immobilization of the protein probe into a surface. The capture region can be an affinity tag as described below or a solid surface such as bead, a slide or an array.

[0080] Figure 6 shows a schematic representation of one of the embodiments of the invention. In this embodiment a protein probe is attached to a capture region, e.g. a magnetic bead. Figure 6 depicts the use of an antibody. However, the embodiment depicted in this figure can utilize any of the protein probes and capture regions described herein. The protein probes (e.g., antibodies) can be attached to a capture region by any suitable method knows in the art including the methods described herein. The target protein is mixed with the protein probe containing the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1222

capture region. The complex of target protein and protein probe is then contacted with a second protein probe attached to a signal oligo via a linker oligo. The complex of target protein and protein probes are purified. In this example, the complex of target protein and antibody is purified using the magnetic bead in the capture antibody. However, in this or any other embodiment described herein, the complex of target protein and protein probes can be purified by any suitable method known in art such as the methods described above. If the capture region is a slide or an array, the complex of target protein and protein probes can be purified by washing off the excess of unbound sample and protein probes. The isolated target protein/protein probes complex is then washed and the signal oligo is eluted. The signal oligo is analyzed using nanoreporters as described below. Methods for eluting the signal oligos are know in the art including the methods described herein. In this embodiment, the proteins and nanoreporters are largely separate, which eliminates concerns about protein stickiness. In some embodiments, the methods depicted in Figure 6 are used to detect and/or quantify a plurality of target proteins. Each target protein will be detected by a probe set comprising a first probe and a second probe as described in Figure 6. Each probe set will have a specific and/or unique signal oligo that can then be associated to the target protein of each probe set.

[0081] In some embodiments, the signal oligo is attached to an affinity tag. The affinity tag in the signal oligo can be used to isolate and/or immobilized the signal oligo. In any of the methods described herein utilizing a signal oligo, the signal oligo can be attached to an affinity tag.

[0082] Figure 7 shows a schematic representation of one of the embodiments of the invention. This embodiment can be used with any of the methods described herein. The diagram is Figure 7 shows antibodies as protein probes, however, this example can be used with any of the protein probes described herein. Figure 7 shows an antibody attached directly or indirectly (e.g. via hybridization through an oligo) to a capture region (e.g. a magnetic bead) and a second antibody attached to a biotinylated signal oligo. However, the embodiment depicted in this figure can utilize any of the capture regions and affinity tags described herein. The target protein is mixed with the protein probes. The complex of target protein and antibodies is purified using the magnetic bead in the capture antibody. However, in this or any other embodiment described herein, the complex of target protein and protein probes can be purified by any suitable method known in art such as the methods described above. If the capture region is a slide or an array, the complex of target protein and protein probes can be purified by washing off the excess of

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1223

unbound sample and protein probes. The isolated target protein/antibody complex is then washed and the signal oligo is eluted by any suitable method known in the art including those described herein. In the embodiment of Figure 7, the signal oligo is purified using oligonucleotide-coupled beads such as Dynabeads®. However, the signal oligo can be purified by any suitable method according to the affinity tag attached to it. The signal oligo is analyzed using nanoreporters as described below. In some embodiments, the methods depicted in Figure 7 are used to detect and/or quantify a plurality of target proteins. Each target protein will be detected by a probe set comprising a first probe and a second probe as described in Figure 7. Each probe set will have a specific and/or unique signal oligo that can then be associated to the target protein of each probe set. The embodiments described in Figure 7 provide the advantage that it requires only two bead purifications. In addition, in this embodiment, proteins and nanoreporters are largely separate, which eliminates concerns about protein stickiness.

[0083] In some embodiments, the signal oligo is generated by ligating two oligos that are in close proximity, e.g., proximity ligation. A diagram of proximity ligation is depicted in Figure 8. In step 1) of Figure 8 probes containing the oligos are designed to bind pairwise to a target protein and to form a signal oligo by ligation when the probes are brought in proximity. Figure 8 shows an embodiment using antibodies as protein probes. However, the method described in Figure 8 can be used with any of the protein probes described herein. The probes containing the oligos can be prepared and purified by any methods known in the art, for example the methods described in Gullberg et al, PNAS 101(22), p 8420-24 (2004). In step 2) of Figure 8, the target protein is then mixed with the probes containing the oligos and the bridging oligos.

[0084] A bridging oligo comprises a polynucleotide sequence. The oligos attached to protein probes comprise a portion that is complementary to the bridging oligo. In some embodiments the complementary portions of the oligos are 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, 50, 60, 70 or more nucleotide bases. In some embodiments the complementary portions of the bridging oligo with each of the oligos attached to the protein probe are 6 to 15 bases, with a total length of bridging oligo is 12-30 bases. In some embodiments, the complementary portions of the oligos are 40 bases. In some embodiments, the complementary portions of the oligos are 30 bases. In some embodiments, the complementary portions of the oligos are 20 bases.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1224

[0085] In step 4) of Figure 8, the components required for probe ligation are added. The oligos in the protein probes can be ligated by any suitable method known in art. Ligation according to the present invention comprises any enzymatic or chemical process wherein an inter-nucleotide linkage is formed between the opposing ends of nucleic acid sequences that are adjacently hybridized to the bridging oligo. Example of enzymes that can be used for ligation include but are not limited to DNA ligase, and RNA ligase such as T4 DNA ligase, T4 RNA ligase, Thermus thermophilus (Tth) ligase, Thermus aquaticus (Taq) DNA ligase, or Pyrococcus furiosus (Pfu) ligase. Chemical ligation can be performed using activating and reducing agents such as carbodiimide, cyanogen bromide (BrCN), imidazole, 1-methylimidazole/carbodiimide/cystamine, N-cyanoimidazole, dithiothreitol (DTT) and ultraviolet light. Also within the scope of the invention are ligation techniques such as gap-filling ligation, including, without limitation, gap-filling OLA and LCR, bridging oligonucleotide ligation, and correction ligation. Descriptions of these techniques can be found, among other places, in U.S. Pat. No. 5,185,243, published European Patent Applications EP 320308 and EP 439182, and PCT Publication Nos. WO 90/01069 and WO 01/57268.

[0086] In step 5) of Figure 8, after ligation, the signal oligo is then released via disulfide reduction, uracil excision, restriction digest, proteinase K, or any other suitable method know in the art. Additionally, the signal oligo can be released by the methods depicted in Figure 8B-8D. Figure 8B, shows an embodiment in which the signal oligo has an affinity tag such as biotin or a sequence. The affinity tag can be used to isolate and/or immobilized the signal oligo as described herein. Figure 8C shows an embodiment in which the bridging oligo has an affinity tag such as biotin or a sequence. The affinity tag can be used to isolate and/or immobilized the signal oligo as described herein. Only the ligated oligo will have enough overlap to remain hybridized to the signal oligo during the isolation and/or immobilization process. Figure 8D shows an embodiment in which the embodiments of Figure 8B and 8C are combined. The signal oligo is analyzed using nanoreporters as described below. In some embodiments, the methods depicted in Figure 8 are used to detect and/or quantify a plurality of target proteins. Each target protein will be detected by a probe set comprising a first probe and a second probe as described in Figure 8. Each probe set will have a specific and/or unique signal oligo that can then be associated to the target protein of each probe set. The embodiments described in Figure 8 have several benefits around sensitivity, minimization of cross-reactivity, and multiplexing. Proximity ligations have

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1225

shown high sensitivity and have the effect of lowering the apparent Kd by essentially decreasing the off-rate.

[0087] In some embodiments utilizing proximity ligation one of the oligos is attached to a nanoreporter. Figure 9 shows a diagram of one of such embodiments.

[0088] In step 1) of Figure 9 probes containing the oligos are designed to bind pairwise to target proteins. One of the oligos in one of the protein probes is attached to a nanoreporter. Figure 9 shows an embodiment using antibodies as protein probes. However, the method described in Figure 9 can be used with any of the protein probes described herein. The probes containing the oligos can be prepared and purified as described above. In step 2) and 3) of Figure 9, the target protein is then mixed with the probes containing the oligos and the bridging oligos. The bridging oligo binds to the oligo in a first protein probe and a portion of the nanoreporter attached to the second protein probe.

[0089] The oligo attached to the first protein probe and the nanoreporter comprise a portion that is complementary to the bridging oligo. In some embodiments the complementary portion is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, 50, 60, 70 or more nucleotide bases. In some embodiments, the complementary portion is 40 bases. In some embodiments, the complementary portion is 30 bases. In some embodiments, the complementary portion is 20 bases. In some embodiments the complementary portions of the bridging oligo with each the oligos attached to the protein probe and the nanorereporter is 6 to 15 bases, with a total length of bridging oligo is 12-30 bases. In step 4) of Figure 9, the components required for probe ligation are then added. The oligo in the first protein probe and the nanoreporter can be ligated by any suitable method known in art as described above. In step 5) of Figure 9, after ligation, the signal oligo can be optionally released via disulfide reduction, uracil excision, restriction digest, proteinase K, or any other suitable method know in the art.

[0090] Additionally, the signal oligo can be released by the methods depicted in Figure 8B-8D. For instance, using the approach described in Figure 8C, a purification step is performed to separate ligated oligos from non-ligated oligos after release of the signal oligo from, for instance, an antibody. This purification step can be performed using magnetic beads or any other method known in the art for the physical separation of proteins. Importantly, if the amount of antibody used is higher than the amount of reporters used, then the resultant excess of unligated oligos may block the hybridization of the reporter to the oligo. As described in Example 7, the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1226

purification step further includes a rinsing step with a buffer solution. Figure 13 demonstrates how various components of an antibody reporter complex are purified and rinsed in a variety of buffer conditions. A preferred rinsing buffer is SSPE; however, other buffers and all concentrations having similar capacities for retaining counts of a reporter complex or a component thereof are encompassed by these methods.

[0091] The signal oligo is analyzed using nanoreporters as described below. In some embodiments, the methods depicted in Figure 9 are used to detect and/or quantify a plurality of target proteins. Each target protein will be detected by a probe set comprising a first probe and a second probe as described in Figure 9. Each probe set will have a specific and/or unique signal oligo that can then be associated to the target protein of each probe set. The embodiments described in Figure 9 take advantage of the decrease in the Koff via proximity ligation. A lower Koff means a lower Kd and the ability to work with lower concentrations of protein probe. This decrease in Kd makes it easier to work in concentrations required for reporters, and thus to contemplate direct detection approaches for multiplex analysis and lower reagent costs. These embodiments do not need a step for hybridization to reporters within the assay. Thus, these assays will be faster and have a shorter time to answer.

[0092] In some embodiments, the signal oligo is analyzed/detected using nanoreporter(s) as described in sections below. In these embodiments, the nanoreporter(s) comprise a portion that is complementary to the signal oligo. In some embodiments the complementary portion is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, 50, 60, 70 or more nucleotide bases. In some embodiments, the complementary portion is 40 bases. In some embodiments, the complementary portion is 30 bases. In some embodiments, the complementary portion is 20 bases. In some embodiments, the complementary portion 15-20 bases.

[0093] In some embodiments, the methods described herein further comprise protein probes containing a nanoreporter. Thus, in some embodiments the invention provides methods for detection and/or quantification of a target protein by binding a protein probe or protein probe set to a target protein, wherein the protein probe or at least one of the protein probes in the probe set contains a nanoreporter.

[0094] Figure 4 shows a schematic diagram of one of the embodiments of the invention. In this embodiment a nanoreporter is attached to one of the antibodies. The methods described in

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1227

Figure 4 can be utilized using any of the protein probes described herein. In some embodiments, the nanoreporter can be directly attached to the protein probe. In other embodiments, the nanoreporter can be attached to a protein probe via hybridization through a linker oligo. Thus, the nanoreporter comprises a portion that is complementary to the linker oligo in the protein probe. In some embodiments the complementary portion is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, 50, 60, 70 or more nucleotide bases. In some embodiments the linker oligo is 15-20 bases. In some embodiments, the complementary portion is 40 bases. In some embodiments, the complementary portion is 30 bases. In some embodiments, the complementary portion is 20 bases. In some embodiments, the complementary portion is 15 bases.

[0095] The hybridization of the nanoreporter to the linker oligo can occur at different temperatures depending of the length of the complementary portion. In some embodiments, the nanoreporter can be hybridized to a linker oligo attached to a protein probe at a temperature in the range of 32°C to 40°C. In some embodiments, the nanoreporter can be hybridized to a linker oligo attached to a protein probe at a temperature of 35°C. In some embodiments, the nanoreporter can be hybridized to a linker oligo attached to a protein probe at a temperature of 37°C. In some embodiments, the nanoreporter can be hybridized to a linker oligo attached to a protein probe at a temperature of 45°C. In some embodiments, the nanoreporter can be hybridized to a linker oligo attached to a protein probe at a temperature of 52-57°C. In some embodiments, the nanoreporter can be hybridized to a linker oligo attached to a protein probe at a temperature of 15-20°C below the melting temperature (Tm) of the complementary portions of the nanoreporter with the linker oligo. One of ordinary skilled in art will understand that the length of the complementary portions of the nanoreporter with the linker oligo and their hybridization temperature will depend on the type of protein probe used. In some embodiments, the protein probe is an antibody and the length of the complementary portions of the nanoreporter with the linker oligo is 15-20 bases, which gives a Tm of about 57°C or 15-20 °C above the ideal antibody temperature of 37°C. Thus in some embodiments, the protein probe is an antibody, the length of the complementary portions of the nanoreporter with the linker oligo is 15-20 bases and the hybridizing temperature is 37°C.

[0096] Figure 4 shows that a complex of target protein and antibodies is formed in which one of the antibodies is bound to biotin and the other antibody has a nanoreporter attached. The

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1228

methods described in Figure 4 can use any affinity tag described herein besides biotin. Purification of the target protein-antibodies complex can be performed by any suitable method known in the art including those described herein. Elution of the nanoreporter can be accomplished by melting off G and F beads, via digestion or any other suitable method known in the art. In the embodiments in which the protein-antibodies complex contain an affinity tag, the complex can be bound to a coverslip, e.g., coated with streptavidin (Optichem®, Accelr8 Technology Corporation). The nanoreporter is analyzed as described below. In some embodiments, the methods depicted in Figure 4 are used to detect and/or quantify a plurality of target proteins. Each target protein will be detected by a probe set comprising a first probe and a second probe as described in Figure 4. Each probe set will have a specific and/or unique nanoreporter that can then be associated to the target protein of each probe set.

[0097] Without intending to be limited to any theory or any specific embodiments, the embodiments of the inventions that utilize a signal oligo present several advantages: (1) these embodiments separate the target proteins and the protein probes from the nanoreporters. Separation of the proteins from the reporters eliminates the potential problems of solubility and stickiness associated with using nanoreporters to measure proteins. Separation of the target proteins from the nanoreporters avoids the Kd mismatch issues between DNA and proteins, allows for the use of ideal concentrations for both to get maximum signal and lowest noise, and allows for the use of low Kd antibodies if needed; (2) the indirect signal oligo approach can be run as a process upstream of the nanoreporter assay described below, thereby taking advantage of an optimized nanoreporter assay; (3) protein probe sets (e.g, antibody pairs) can be used in their normal configuration if needed, e.g., capture antibody on surface (on a magnetic bead for example), and detection antibody in solution. Some antibodies work best in this configuration; (4) with these embodiments problems associated with the protein probes coming off the target (Koff rate) are minimized, e.g., antibodies only have to stay bound to the target during binding and purification on the beads. This allows for use of a large range of antibodies including antibodies with lower binding affinity; and (5) proteins can be read in the same lane as nucleic acids, e.g., RNA or DNA. The sample is first split: part is run through the protein detection embodiments described herein (lyse cells with detergent then bind and purify as described herein), and part is split off and processed as nucleic acid samples (cells are lysed with GITC). The samples are then recombined and analyzed using nanoreporters as described below,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1229

potentially in the same lane. Measurement of both nucleic acids (e.g., RNA) and proteins in the same lane will minimize measurement differences, make protein and nucleic acid expression data more comparable, and eliminate the need for multiple measurement methods to get the required data.

[0098] In some embodiments, the methods described herein provide for the measurement of nucleic acids, e.g., RNA or DNA, in combination with the measurement of proteins.

[0099] Any of the embodiments described herein can be used in the detection of multiple target proteins. In some embodiments, the invention provides methods comprising protein probes for the analysis of target proteins. In some embodiments, the invention provides a protein probe population for use in a multiplexed assay. Each protein probe in the population is specific for a target molecule. The binding of the target proteins to the protein probes is then detected using nanoreporters. Each nanoreporter comprises a unique label code that can be associated to a specific target molecule as described below.

[00100] In some embodiments, the detection of the nanoreporters as described below is digital in nature in that one molecule at a time is counted. While fluorescence is used to read the code, the signals are high and the spot is either present of not, thus the digital detection. Using digital detection rather than an analogue fluorescent signal used to quantify signal leads to more accurate quantification. Thus the methods described herein allows for multiplexing to levels beyond currently possible, for more accurate quantification, and possibly higher sensitivity.

Nanoreporters

[00101] A nanoreporter which provides a code of signals (the nanoreporter label code) associated with a specific target. In some embodiments, upon binding of the nanoreporter to a signal oligo or a linker oligo associated with a protein probe, the nanoreporter code identifies the signal oligo or the protein probe to which the nanoreporter is bound. Thus, in some embodiments the nanoreporters of the invention comprise two main portions: (i) a sequence specific for a signal oligo-specific or a linker oligo associated with a protein probe; and (ii) a labeled nanoreporter. In some embodiments, the nanoreporters are directly attached to a protein probe.

[00102] Nanoreporters are modular structures. In some embodiments, the nanoreporter comprises a plurality of different detectable molecules. In some embodiments, a labeled

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1230

nanoreporter is a molecular moiety containing certain basic elements: (i) a plurality of label attachment regions attached in linear combination, and (ii) complementary polynucleotide sequences attached to the label attachment regions of the backbone. In some embodiments, the labeled nanoreporter comprises 2, 3, 4, 5, 6, 7, 8, 9, 10 or more unique label attachment regions attached in a linear combination, and complementary polynucleotide sequences attached to the label attachment regions of the backbone. In some embodiments, the labeled nanoreporter comprises 3 or more label attachment regions attached in linear combination, and complementary polynucleotide sequences attached to the label attachment regions of the backbone. The term label attachment region includes a region of defined polynucleotide sequence within a given backbone that may serve as an individual attachment point for a detectable molecule.

[00103] The plurality of label attachment regions attached in linear combination can comprise uniquely designed sequences. In addition, the plurality of label attachment regions attached in linear combination in the nanoreporters can comprise at least one template, for example but not limited to, at least one nucleic acid sequence, such as at least part of a linear or linearizable viral genome, such as the genomes of adenovirus, hepatitis virus, herpes virus, rotavirus, and the like, or bacteriophages such as lambda, M13, ϕX-174, T-series bacteriophages, and the like, including derivatives thereof comprising cloning cassettes, polylinkers, and the like; plasmids, such as pBR322 and pUC series plasmids, etc., including derivatives thereof comprising cloning cassettes, polylinkers, and the like; synthetic templates; templates comprising artificial sequences; and the like. The skilled artisan will understand that virtually any piece of nucleic acid can serve as a template for fabricating a nanoreporter provided that it is large enough to include at least two label attachment regions, or it can be combined with at least one other nucleic acid sequence so that the combined sequence is large enough to include at least two label attachment regions.

[00104] In some embodiments, the labeled nanoreporter also comprises a backbone containing a constant region. The constant region can be directly or indirectly attached to the nanoreporter. Thus, the constant region can covalently attached to a nanoreporter or the constant region can be bound to the nanoreporter later in the assay. The term constant region includes tandemly-repeated sequences of about 10 to about 25 nucleotides. The constant region can be attached at either the 5' region or the 3' region of a nanoreporter, and may be utilized for capture

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1231

and immobilization of a nanoreporter for imaging or detection, such as by attaching to a solid substrate a sequence that is complementary to the constant region.

[00105] The elements of a nanoreporter can be found in a single molecular moiety (a singular nanoreporter), or two distinct molecular moieties (a dual nanoreporter). Each molecular moiety may be composed of one molecule or more than one molecule attached to one another by covalent or non-covalent means. In some embodiments, each component of a dual nanoreporter has a signal oligo-specific sequence that binds to a different site on the same signal oligo molecule. When using a dual nanoreporter system one of the nanoreporter probes may be unlabeled. In some embodiments, the unlabeled nanoreporter probe may comprise a capture region. In some embodiments, the unlabeled nanoreporter probe may comprise a signal oligo-specific region and a backbone that may be single stranded. In some embodiments, the unlabeled nanoreporter probe may comprise a signal oligo-specific region and a backbone that may be double stranded.

[00106] The complementary polynucleotide sequences attached to a nanoreporter backbone serve to attach detectable molecules, or label monomers, to the nanoreporter backbone. The complementary polynucleotide sequences may be directly labeled, for example, by covalent incorporation of one or more detectable molecules into the complementary polynucleotide sequence. Alternatively, the complementary polynucleotide sequences may be indirectly labeled, such as by incorporation of biotin or other molecule capable of a specific ligand interaction into the complementary polynucleotide sequence. In such instances, the ligand (e.g., streptavidin in the case of biotin incorporation into the complementary polynucleotide sequence) may be covalently attached to the detectable molecule. Where the detectable molecules attached to a label attachment region are not directly incorporated into the complementary polynucleotide sequence, this sequence serves as a bridge between the detectable molecule and the label attachment region, and may be referred to as a bridging molecule, e.g., a bridging nucleic acid.

[00107] In some embodiments, the invention uses the nanoreporters described in US patent 7,473,767; US applications No. 10/542,458; 12/324,357; 11/645,270 and 12/541,131, incorporated herein by reference in their entireties.

[00108] The nucleic-acid based nanoreporter, nanoreporter-signal oligo complexes, or nanoreporter-protein probe complexes of the present invention comprise nucleic acids, which may be affinity-purified or immobilized using a nucleic acid, such as an oligonucleotide, that is

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1232

complementary to the constant region of the nanoreporter. As noted above, in some embodiments the nanoreporters comprise at least one constant region, which may serve as an affinity tag for purification and/or for immobilization (for example to a solid surface). The constant region typically comprises two or more tandemly-repeated regions of repeat nucleotides, such as a series of 15-base repeats. In such exemplary embodiments, the nanoreporter, whether complexed to a signal oligo, a target molecule or otherwise, can be purified or immobilized by an affinity reagent coated with a 15-base oligonucleotide which is the reverse complement of the repeat unit.

[00109] Nanoreporters, nanoreporter-signal oligo complexes, or nanoreporter-protein probe complexes can be purified in two or more affinity selection steps. For example, in the embodiments in which the nanoreporter is attached to a protein probe, the nanoreporter can comprise an affinity tag. In other embodiments when a signal oligo and dual nanoreporters are used, one nanoreporter probe can comprise a first affinity tag and the other nanoreporter probe can comprise a second (different) affinity tag. The nanoreporter probes are mixed with the signal oligos, and complexes comprising the two probes of the dual nanoreporters are separated from unbound materials (e.g., the signal oligo or the individual probes of the nanoreporter) by affinity purification against one or both individual affinity tags. In the first step, the mixture can be bound to an affinity reagent for the first affinity tag, so that only probes comprising the first affinity tag and the desired complexes are purified. The bound materials are released from the first affinity reagent and optionally bound to an affinity reagent for the second affinity tag, allowing the separation of complexes from nanoreporter probes comprising the first affinity tag. At this point only full complexes would be bound. The complexes are finally released from the affinity reagent for the second affinity tag and then preferably stretched and imaged. The affinity reagent can be any solid surface coated with a binding partner for the affinity tag, such as a column, bead (e.g., latex or magnetic bead) or slide coated with the binding partner. Immobilizing and stretching nanoreporters using affinity reagents is fully described in U.S. Provisional Application no. 60/753,816 by Sean M. Ferree and Dwayne L. Dunaway, entitled "Compositions Comprising Oriented, Immobilized Macromolecules and Methods for Their Preparation," filed on December 23, 2005, and US patent 7,473,767; US applications No. 10/542,458; 12/324,357; 11/645,270 and 12/541,131, incorporated herein by reference in their entirety.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1233

[00110] The sequence of signals provided by the label monomers associated with the various label attachment regions of the backbone of a given nanoreporter allows for the unique identification of the nanoreporter. For example, when using fluorescent labels, a nanoreporter having a unique identity or unique spectral signature is associated with a signal oligo-specific sequence or a protein probe that recognizes a specific target molecule or a portion thereof. Detection of the nanoreporter signal, such as the spectral code of a fluorescently labeled nanoreporter, associated with the nanoreporter allows detection of the presence of the target molecule in the mixture (qualitative analysis). Counting all the label monomers associated with a given spectral code or signature allows the counting of all the molecules in the mixture associated with the signal oligo -specific sequence or the protein probe coupled to the nanoreporter (quantitative analysis). In the embodiments where a signal oligo is used, the signal oligos then can be correlated to the target molecule via the binding of target molecule to the protein probe associated with the signal oligo. Nanoreporters are thus useful for the diagnosis or prognosis of different biological states (e.g., disease vs. healthy) by quantitative analysis of known biological markers.

[00111] Moreover, the exquisite sensitivity of single molecule detection and quantification provided by the nanoreporters of the invention allows for the identification of new diagnostic and prognostic markers, including those whose fluctuations among the different biological states is too slight detect a correlation with a particular biological state using traditional molecular methods. The sensitivity of nanoreporter-based molecular detection permits detailed pharmacokinetic analysis of therapeutic and diagnostic agents in small biological samples.

[00112] Nanoreporters' syntheses can be performed by any suitable methods known in the art. Examples of nanoreporters' syntheses are described in US patent 7,473,767; US applications No. 10/542,458; 12/324,357; 11/645,270 and 12/541,131, incorporated herein by reference in their entireties.

[00113] In one embodiment, the invention provides a nanoreporter further comprising an affinity tag attached to the nanoreporter backbone, such that attachment of the affinity tag to a support allows backbone stretching and resolution of signals provided by label monomers corresponding to different label attachment regions on the backbone. Nanoreporter stretching may involve any stretching means known in the art including but not limited to, means involving physical, hydrodynamic or electrical means. The affinity tag may comprise a constant region.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1234

[00114] The uniqueness of each nanoreporter probe in a population of probe allows for the multiplexed analysis of a plurality of target molecules. For example, in some embodiments, each nanoreporter probe can contain contains six label attachment regions, where each label attachment region of each backbone is different from the other label attachment regions in that same backbone. If the label attachment regions are going to be labeled with one of four colors and there are 24 possible unique sequences for the label attachment regions and each label attachment region is assigned a specific color, each label attachment region in each backbone will consist of one of four sequences. There will be 4096 possible nanoreporters in this example. The number of possible nanoreporters can be increased, for example, by increasing the number of colors, increasing the number of unique sequences for the label attachment regions and/or increasing the number of label attachment regions per backbone. Likewise the number of possible nanoreporters can be decreased by decreasing the number of colors, decreasing the number of unique sequences for the label attachment regions and/or decreasing the number of label attachment regions per backbone.

[00115] In certain embodiments, the methods of detection are performed in multiplex assays, whereby a plurality of target molecules is detected in the same assay (a single reaction mixture). In a preferred embodiment, the assay is a hybridization assay in which the plurality of target molecules is detected simultaneously. In certain embodiments, the plurality of target molecules detected in the same assay is, at least 2, at least 5 different target molecules, at least 10 different target molecules, at least 20 different target molecules, at least 50 different target molecules, at least 75 different target molecules, at least 100 different target molecules, at least 200 different target molecules, at least 500 different target molecules, or at least 750 different target molecules, or at least 1000 different target molecules. In other embodiments, the plurality of target molecules detected in the same assay is up to 50 different target molecules, up to 100 different target molecules, up to 150 different target molecules, up to 200 different target molecules, up to 300 different target molecules, up to 500 different target molecules, up to 750 different target molecules, up to 1000 different target molecules, up to 2000 target molecules, or up to 5000 target molecules. In yet other embodiments, the plurality of target molecules detected is any range in between the foregoing numbers of different target molecules, such as, but not limited to, from 20 to 50 different target molecules, from 50 to 200 different target molecules,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1235

from 100 to 1000 different target molecules, from 500 to 5000 different target molecules, and so on and so forth.

[00116] In addition to the qualitative analytical capabilities provided by the nanoreporters of the invention and the analytical techniques based thereon, the nanoreporters of the invention are uniquely suitable for conducting quantitative analyses. By providing a one to one binding between the nanoreporters (whether singular or dual nanoreporters) of the invention and their target molecules in a biomolecular sample, all or a representative portion of the target molecules present in the sample can be identified and counted. This individual counting of the various molecular species provides an accurate and direct method for determining the absolute or relative concentration of the target molecule in the biomolecular sample. Moreover, the ability to address each molecule in a mixture individually leverages benefits of miniaturization including high sensitivity, minimal sample quantity requirements, high reaction rates which are afforded by solution phase kinetics in a small volume, and ultimately very low reagent costs.

Detectable Molecules or Label Monomers

[00117] The nanoreporters of the present invention can be labeled with any of a variety of label monomers, such as a radioisotope, fluorochrome, dye, enzyme, nanoparticle, chemiluminescent marker, biotin, or other monomer known in the art that can be detected directly (e.g., by light emission) or indirectly (e.g., by binding of a fluorescently-labeled antibody). Generally, one or more of the label attachment regions in the nanoreporter is labeled with one or more label monomers, and the signals provided by the label monomers attached to the label attachment regions of a nanoreporter constitute a detectable code that identifies the target to which the target-specific region of the nanoreporter binds. In certain embodiments, the lack of a given signal from the label attachment region (e.g., a dark spot) can also constitute part of the nanoreporter code.

[00118] Example of label monomers that can be used with the nanoreporters described herein and methods to incorporate the labels monomers into the nanoreporters are described in US patent 7,473,767; US applications No. 10/542,458; 12/324,357; 11/645,270 and 12/541,131, incorporated herein by reference in their entireties.

Affinity Tags

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1236

[00119] A variety of affinity tags known in the art may be used, e.g., to purify and/or immobilize nanoreporters. In some embodiments, a biotin anchor is attached to the nanoreporter, allowing immobilization of the nanoreporter on a streptavidin coated slide.

[00120] In some embodiments, a labeled nanoreporter will contain an affinity tag at each end, A1 and A2. The labeled nanoreporter can be immobilized to a surface through the binding of A1 to an immobilized affinity partner. In the absence of an affinity binding partner for A2, the A2 end of the nanoreporter remains in solution, but in the presence of an affinity binding partner (A2'), the A2 end of the nanoreporter is also immobilized. In some embodiments, a labeled nanoreporter will contain a single affinity tag, A1. Another affinity tag, A2, can be attached to the nanoreporter by direct binding of the nanoreporter to a molecule containing A2 (e.g., if the nanoreporter is or comprises a nucleic acid, it can hybridize directly with another nucleic acid to which A2 is attached). Alternatively, either affinity tag can be attached to the labeled nanoreporter via a bridging molecule, such as the bridging nucleic acid. In some embodiments, upon immobilization of A1, the nanoreporter can be stretched, or "elongated", for example by electrostretching, for separation of the label attachment regions in a manner that permits detection of the nanoreporter code. Optionally, while the nanoreporter is in an elongated state, A2 is introduced and binds the end of the nanoreporter that is complementary to A2 down to the surface.

[00121] In some embodiments, an affinity tag is attached to a protein probe, e.g., to purify and/or immobilize the protein probe.

[00122] An affinity tag can be used for attachment to beads or other matrixes for a variety of useful applications including but not limited to purification.

[00123] Examples of affinity tags and methods of making and/or attaching them to the nanoreporters described herein are described in US patent 7,473,767; US applications No. 10/542,458; 12/324,357; 11/645,270 and 12/541,131, incorporated herein by reference in their entireties.

Biomolecular Samples

[00124] The protein probe and nanoreporter systems of the invention can be used to detect target proteins in any biomolecular sample. As will be appreciated by those in the art, the sample may comprise any number of things, including, but not limited to: biological samples, such as cells (including both primary cells and cultured cell lines), cell lysates, or extracts,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1237

tissues and tissue extracts; bodily fluids (including, but not limited to, blood, urine, serum, lymph, bile, cerebrospinal fluid, interstitial fluid, aqueous or vitreous humor, colostrum, sputum, amniotic fluid, saliva, anal and vaginal secretions, perspiration and semen, a transudate, an exudate (e.g., fluid obtained from an abscess or any other site of infection or inflammation) or fluid obtained from a joint (e.g., a normal joint or a joint affected by disease such as rheumatoid arthritis, osteoarthritis, gout or septic arthritis) of virtually any organism, with mammalian samples being preferred and human samples being particularly preferred; environmental samples (including, but not limited to, air, agricultural, water and soil samples); biological warfare agent samples; research samples including extracellular fluids, extracellular supernatants from cell cultures, inclusion bodies in bacteria, cellular compartments, cellular periplasm, mitochondria compartment, etc.

[00125] The biomolecular samples can be indirectly derived from biological specimens. For example, where the target protein of interest is a kinase the biomolecular sample of the invention can be a sample containing isolated proteins from a cell lysate. In another example, the biomolecular sample of the invention is generated by subjecting a biological specimen to fractionation, e.g., size fractionation or membrane fractionation.

[00126] The biomolecular samples of the invention may be either native, e.g., not subject to manipulation or treatment, or treated, which can include any number of treatments, including exposure to candidate agents including drugs, genetic engineering (e.g., the addition or deletion of a gene), etc.

[00127] Biomolecular samples may also include environmental samples, such as those containing bacteria or other organisms, such as diatoms, dinoflagellates, algae, among others, such as in certain marine or earth-based samples.

Detection of Nanoreporters

[00128] Nanoreporters are detected by any means available in the art that is capable of detecting the specific signals on a given nanoreporter. Where the nanoreporter is fluorescently labeled, suitable consideration of appropriate excitation sources may be investigated. Possible sources may include but are not limited to arc lamp, xenon lamp, lasers, light emitting diodes or some combination thereof. The appropriate excitation source is used in conjunction with an appropriate optical detection system, for example an inverted fluorescent microscope, an epi-

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1238

fluorescent microscope or a confocal microscope. Preferably, a microscope is used that can allow for detection with enough spatial resolution to determine the sequence of the spots on the nanoreporter. For example in one embodiment an image of a dual nanoreporter hybridized to a target molecule can be obtained. If for example, the nanoreporters are labeled with three different colors, Alexa 488, Cy3 and Alexa 647 (labeled 1, 2 and 3, respectively). Colors 1, 2 and 3 are each acquired in different channels and the first and second registers, which can be seen as rows of spots, are shifted up by several pixels to be able to show each register individually.

[00129] Examples of methods for detection of nanoreporters that can be used in the methods of the invention are described in US patent 7,473,767 entitled "Methods for detection and quantification of analytes in complex mixtures", US patent publication no. 2007/0166708 entitled "Methods for detection and quantification of analytes in complex mixtures", US application number 11/645,270 entitled "Compositions comprising oriented, immobilized macromolecules and methods for their preparation", PCT application no US06/049274 entitled "Nanoreporters and methods of manufacturing and use thereof", and US provisional application 60/088,988 entitled "Stable nanoreporter", all of which are incorporated by reference herein in its entirety.

Applications for Protein Detection via Nanoreporter Technology

[00130] The compositions and methods of the invention can be used for diagnostic, prognostic, therapeutic, patient stratification, drug development, treatment selection and screening purposes. The present invention provides the advantage that many different target proteins can be analyzed at one time from a single biomolecular sample using the methods of the invention. This allows, for example, for several diagnostic tests to be performed on one sample.

[00131] The composition and methods of the invention can be used in proteomics. The methods described herein will typically provide an answer rapidly which is very desirable for this application. The methods and composition described herein can be used in the process of finding biomarkers that may be used for diagnostics or prognostics and as indicators of health and disease. The methods and composition described herein can be used to screen for drugs, e.g., drug development, selection of treatment, determination of treatment efficacy and/or identify targets for pharmaceutical development. The ability to test protein expression on screening assays involving drugs is very important because proteins are the final gene product in the body. In some embodiments, the methods and compositions described herein will measure

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1239

both protein and gene expression simultaneously which will provide the most information regarding the particular screening being performed.

[00132] The present methods can be applied to the analysis of biomolecular samples obtained or derived from a patient so as to determine whether a diseased cell type is present in the sample, the stage of the disease, the prognosis for the patient, the ability to the patient to respond to a particular treatment, or the best treatment for the patient. The present methods can also be applied to identified biomarkers for a particular disease

[00133] In some embodiments, the methods described herein are used in the diagnosis of a condition. As used herein the term "diagnose" or "diagnosis" of a condition includes predicting or diagnosing the condition, determining predisposition to the condition, monitoring treatment of the condition, diagnosing a therapeutic response of the disease, and prognosis of the condition, condition progression, and response to particular treatment of the condition. For example, a blood sample can be assayed according to any of the methods described herein to determine the presence and/or quantity of markers of a disease or malignant cell type in the sample, thereby diagnosing or staging the a disease or a cancer.

[00134] In some embodiments, the methods and composition described herein are used for the diagnosis and prognosis of a condition.

[00135] Numerous immunologic, proliferative and malignant diseases and disorders are especially amenable to the methods described herein. Immunologic diseases and disorders include allergic diseases and disorders, disorders of immune function, and autoimmune diseases and conditions. Allergic diseases and disorders include but are not limited to allergic rhinitis, allergic conjunctivitis, allergic asthma, atopic eczema, atopic dermatitis, and food allergy. Immunodeficiencies include but are not limited to severe combined immunodeficiency (SCID), hypereosinophilic syndrome, chronic granulomatous disease, leukocyte adhesion deficiency I and II, hyper IgE syndrome, Chediak Higashi, neutrophilias, neutropenias, aplasias, Agammaglobulinemia, hyper-IgM syndromes, DiGeorge/Velocardial-facial syndromes and Interferon gamma-TH1 pathway defects. Autoimmune and immune dysregulation disorders include but are not limited to rheumatoid arthritis, diabetes, systemic lupus erythematosus, Graves' disease, Graves ophthalmopathy, Crohn's disease, multiple sclerosis, psoriasis, systemic sclerosis, goiter and struma lymphomatosa (Hashimoto's thyroiditis, lymphadenoid goiter), alopecia aerata, autoimmune myocarditis, lichen sclerosis, autoimmune uveitis, Addison's

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1240

disease, atrophic gastritis, myasthenia gravis, idiopathic thrombocytopenic purpura, hemolytic anemia, primary biliary cirrhosis, Wegener's granulomatosis, polyarteritis nodosa, and inflammatory bowel disease, allograft rejection and tissue destructive from allergic reactions to infectious microorganisms or to environmental antigens.

[00136] Proliferative diseases and disorders that may be evaluated by the methods of the invention include, but are not limited to, hemangiomatosis in newborns; secondary progressive multiple sclerosis; chronic progressive myelodegenerative disease; neurofibromatosis; ganglioneuromatosis; keloid formation; Paget's Disease of the bone; fibrocystic disease (e.g., of the breast or uterus); sarcoidosis; Peronies and Duputren's fibrosis, cirrhosis, atherosclerosis and vascular restenosis.

[00137] Malignant diseases and disorders that may be evaluated by the methods of the invention include both hematologic malignancies and solid tumors.

[00138] Hematologic malignancies are especially amenable to the methods of the invention when the sample is a blood sample, because such malignancies involve changes in blood-borne cells. Such malignancies include non-Hodgkin’s lymphoma, Hodgkin’s lymphoma, non-B cell lymphomas, and other lymphomas, acute or chronic leukemias, polycythemias, thrombocythemias, multiple myeloma, myelodysplastic disorders, myeloproliferative disorders, myelofibroses, atypical immune lymphoproliferations and plasma cell disorders.

[00139] Plasma cell disorders that may be evaluated by the methods of the invention include multiple myeloma, amyloidosis and Waldenstrom’s macroglobulinemia.

[00140] Example of solid tumors include, but are not limited to, colon cancer, breast cancer, lung cancer, prostate cancer, brain tumors, central nervous system tumors, bladder tumors, melanomas, liver cancer, osteosarcoma and other bone cancers, testicular and ovarian carcinomas, head and neck tumors, and cervical neoplasms.

[00141] The methods described herein can be used to diagnose pathogen infections, for example infections by intracellular bacteria and viruses, by determining the presence and/or quantity of markers of bacterium or virus, respectively, in the sample.

[00142] A wide variety of infectious diseases can be detected by the process of the present invention. Typically, these are caused by bacterial, viral, parasite, and fungal infectious agents. The resistance of various infectious agents to drugs can also be determined using the present invention.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1241

[00143] Bacterial infectious agents which can be detected by the present invention include Escherichia coli, Salmonella, Shigella, Klebsiella, Pseudomonas, Listeria monocytogenes, Mycobacterium tuberculosis, Mycobacterium aviumintracellulare, Yersinia, Francisella, Pasteurella, Brucella, Clostridia, Bordetella pertussis, Bacteroides, Staphylococcus aureus, Streptococcus pneumonia, B-Hemolytic strep., Corynebacteria, Legionella, Mycoplasma, Ureaplasma, Chlamydia, Neisseria gonorrhea, Neisseria meningitides, Hemophilus influenza, Enterococcus faecalis, Proteus vulgaris, Proteus mirabilis, Helicobacter pylori, Treponema palladium, Borrelia burgdorferi, Borrelia recurrentis, Rickettsial pathogens, Nocardia, and Acitnomycetes.

[00144] Fungal infectious agents which can be detected by the present invention include Cryptococcus neoformans, Blastomyces dermatitidis, Histoplasma capsulatum, Coccidioides immitis, Paracoccidioides brasiliensis, Candida albicans, Aspergillus fumigautus, Phycomycetes (Rhizopus), Sporothrix schenckii, Chromomycosis, and Maduromycosis.

[00145] Viral infectious agents which can be detected by the present invention include human immunodeficiency virus, human T-cell lymphocytotrophic virus, hepatitis viruses (e.g., Hepatitis B Virus and Hepatitis C Virus), Epstein-Barr Virus, cytomegalovirus, human papillomaviruses, orthomyxo viruses, paramyxo viruses, adenoviruses, corona viruses, rhabdo viruses, polio viruses, toga viruses, bunya viruses, arena viruses, rubella viruses, and reo viruses.

[00146] Parasitic agents which can be detected by the present invention include Plasmodium falciparum, Plasmodium malaria, Plasmodium vivax, Plasmodium ovale, Onchoverva volvulus, Leishmania, Trypanosoma spp., Schistosoma spp., Entamoeba histolytica, Cryptosporidum, Giardia spp., Trichimonas spp., Balatidium coli, Wuchereria bancrofti, Toxoplasma spp., Enterobius vermicularis, Ascaris lumbricoides, Trichuris trichiura, Dracunculus medinesis, trematodes, Diphyllobothrium latum, Taenia spp., Pneumocystis carinii, and Necator americanis.

[00147] The present invention is also useful for detection of drug resistance by infectious agents. For example, vancomycin-resistant Enterococcus faecium, methicillin-resistant Staphylococcus aureus, penicillin-resistant Streptococcus pneumoniae, multi-drug resistant Mycobacterium tuberculosis, and AZT-resistant human immunodeficiency virus can all be identified with the present invention

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1242

[00148] Thus, the target molecules detected using the compositions and methods of the invention can be either patient markers (such as a cancer marker) or markers of infection with a foreign agent, such as bacterial or viral markers.

[00149] Because of the quantitative nature of nanoreporters, the compositions and methods of the invention can be used to quantitate target protein whose abundance is indicative of a biological state or disease condition, for example, blood markers that are upregulated or downregulated as a result of a disease state.

[00150] In some embodiments, the methods and compositions of the present invention can be used for cytokine detection. The low sensitivity of the methods described herein would be helpful for early detection of cytokines, e.g., as biomarkers of a condition, diagnosis or prognosis of a disease such as cancer, and the identification of subclinical conditions.

Kits

[00151] The invention further provides kits comprising one or more components of the invention. The kits can comprise, for example, one or more protein probe sets and/or one or more nanoreporters. The kits can be used for any purpose apparent to those of skill in the art, including those described above.

[00152] In certain embodiments, the present invention also provides kits useful for the extension and selective immobilization of nanoreporters. The kits can comprise a substrate for immobilization and one or more binding partners to facilitate extension or immobilization of a nanoreporter. The binding partners could, in certain embodiments, comprise a moiety useful for extension of the nanoreporter in an appropriate force. In certain embodiments, the binding partners could facilitate immobilization or selective immobilization of the nanoreporter to the surface. In further embodiments, the kits could comprise a nanoreporter for extension and immobilization. In further embodiments, the kits could comprise a device capable of extending the nanoreporter.

[00153] The kits can contain a population of protein probes and/or nanoreporters as described herein.

[00154] The kits can contain pre-labeled nanoreporters, or unlabeled nanoreporters with one or more components for labeling the nanoreporters. Moreover, the nanoreporters provided in a kit

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1243

may or may not have target-specific sequences pre-attached. In one embodiment, the target sequences are provided in the kit unattached to the nanoreporter backbone.

[00155] The kits can comprise other reagents such as signal oligos, linker oligos and bridging oligos. In some embodiments, the kits can separate the protein probe pairs into different premixes.

[00156] The kits can include other reagents as well, for example, buffers for performing hybridization reactions, linkers, restriction endonucleases, and DNA I ligases.

[00157] The kits also will include instructions for using the components of the kit, and/or for making and/or using the labeled nanoreporters.

[00158] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

EXAMPLES

Example 1. Detection of proteins using indirect measurements - Sandwich assay in solution

[00159] A diagram of the protocol for this Example is depicted in Figure 1. In this Example the assay is set up to separate the binding of the protein target from the hybridization of the reporters to eliminate issues with the mismatch between low binding affinity of protein probes and the working concentration of reporters.

Label of detection antibody with oligo and purification

[00160] The linker oligos were attached to the antibody by random amine attachment using a 10 to 1 ratio of linker oligos to antibodies. Briefly, bifunctional crosslinker Sulfosuccinimidyl 4-[p-maleimidophenyl]butyrate (SMPB) (ThermoFisher, Inc., Waltham, MA) was coupled to anti-IL2 antibody A then reacted to thiolated oligo to crosslinker at an antibody:oligo ratio of 1:3 at room temperature. The antibody A linked to SMPB was purified by running the mixture thru Zeba column 2X, 1000G, (ThermoFisher, Inc., Waltham, MA) and the yield is determined.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1244

[00161] To couple oligo to SMPB linked IL-2 antibody A the oligo was added to purified antibody linked to SMPB at 4°C.

[00162] The oligo linked to IL-2 antibody A the antibody-SMPB-oligo and PBE were added to Pall Nanosep® Centrifugal Device with Omega Membrane, (MWCO 100 kDa, Sigma-Aldrich, Inc., St. Louis, MO) was washed and centrifuged.

Hybridization of signal oligo to detection antibody

[00163] The signal oligo was preannealed to the oligo linked to IL-2 antibody A by adding oligo linked Il2 antibody A and signal oligo in a ratio of 3:2, signal oligo:Antibody ratio Other ratios are contemplated.

Formation of target and antibodies complex

[00164] IL-2 antibody A annealed to signal oligo was mixed with biotinylated antibody B (BAF202,R&D systems, Inc., Minneapolis, MN) at approximately $1x10^{-15}$ to $1x10^{-8}$ M, and blockers (salmon sperm), with room to add target solution. The target protein IL-2 was added to the desired dilution ($<1x10^{-8}$M). Antibodies were at 10X concentration over an estimated Kd of 10^{0-15} to 10^{0-8} M. The mixture was incubated.

[00165] The complex of target protein and antibodies was purified using Streptavidin-coupled Dynabeads® (Invitrogen) according to protocol.

Elution of Signal Oligo

[00166] The isolated target protein/antibody complex was washed and the signal oligo was eluted at greater than 45 ^{o}C using 0.1X SSPE for 10 to 15 minutes. Shorter and longer periods are contemplated *Detection of Signal Oligo*

[00167] Detection of the signal oligo in each sample was carried out using a dual nanoreporter system having both labeled nanoreporter probes and unlabeled nanoreporter probes. The signal oligos from each sample were hybridized with final concentrations of the hybridization reagents as follows: unlabeled, biotinylated probe labeled reporter probe, 5X SSPE (pH 7.5), 5X Denhardt's reagent (Sigma), sheared salmon sperm DNA (Sigma), and detergent. Reagents were mixed and incubated in a thermocycler block with a heated lid for 16 hours.

Post-hybridization purification

[00168] To remove unhybridized reporters, reactions were purified over magnetic beads (InvitrogenTM) coupled to oligonucleotides complementary to the 3'-repeat sequence contained on every biotinylated probe. Reactions were first diluted SSPE in 0.1% detergent mixture/TE and

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1245

allowed to bind to beads at greater than 20°C with continuous rotation. The beads were washed three times in SSPE and detergent and the hybridized complexes eluted in of 0.1X SSPE/0.1% /detergent mixture for 15 minutes at 45°C. After elution, samples were purified a second time to remove excess biotinylated probes by binding to magnetic beads coupled to oligonucleotides complementary to the 5'-repeat sequence contained on every reporter probe. The elutions from the anti-3'-repeat beads were brought to a final concentration of 1X SSPE and bound for 15 minutes at 22.5°C with rotation. Beads were washed as above and eluted in of 0.1X SSPE/0.1% /detergent mixture at greater than 40°C. The doubly-purified samples were then prepared for capture as described below.

NanoString reporter capture, stretching, and imaging

[00169] A solution of a custom-formulation of Tetraspeck fluorescent microspheres (InvitrogenTM) was added to each sample. Samples were loaded into a NanoString fluidic device processed and imaged. *Results:*

[00170] Results are shown in Figure 2. The results of this experiment showed that IL-2 was detected by the assay described herein (Figure 2). This experiment showed a sensitivity of approximately 110^{-11} to 110^{-10} M. The efficiency of detection is the slope shown in the plot in Figure 3. The efficiency observed was probably due to the binding affinity of the antibody to which the signal oligo was attached. This antibody seems to have a Kd of approximately $1.3x10^{-7}$. It is expected that the efficiency can be increased 100X by replacing this antibody with an antibody having a Kd of approximately 10^{-9}.

[00171] It is expected that with improvements in efficiency and modest improvements in background, sensitivity should reach levels of $1x10^{-13}$M but further improvements are possible with continued reduction of background

[00172] This technology allows for multiplexing to levels beyond currently possible, has the potential to allow for more accurate quantification, and possibly higher sensitivity though only further development will prove the last point.

Example 2. Detection of proteins using direct measurement – Solution Triparitide – Protein Probe attached to reporter

[00173] A diagram of the protocol for this Example is depicted in Figure 4. In this Example the nanoreporter is attached to one of the antibodies. Preparation of the antibodies and binding to the sample happens similar to the protocol described in Example 1. In the assay described in

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1246

Figure 4 a complex of target protein and antibodies is formed in which one of the antibodies is bound to biotin and the other antibody has a nanoreporter attached.

[00174] This approach would work best with medium to low Kd antibodies, i.e., strong binding affinity. Without intending to be limited to any theory, The Kd (dissociation constant) of an antibody is usually much higher than the working concentrations of nanoreporters. In order to work with antibodies attached to nanoreporters the Kd needs to be ten times lower than the working concentration of the reporters to insure >90% binding of target to probe. Figure 5 shows a calculation of the ideal Kd for the probes needed in this assay. Figure 5 shows the fraction of target bound vs. the Kd of the nanoreporter probe and the protein probe. Figure 5 shows that the ideal Kd for the protein probes would be approximately $1.0x10^{-15}$ to $1.0x10^{-10}$. This Kd would allow for concentrations of multiplexed reporters below the reporter entanglement threshold of $1x10^{-7}$M.

Attachment of Nanoreporter to Antibody

[00175] One approach is to attach the antibody on the reporter before the binding of the antibody to the target protein. This approach requires very strong binding affinity (Kd) antibodies to allow for concentrations of multiplexed reporters significantly below the reporter entanglement threshold of $1x10^{-7}$M.

[00176] A linker oligo is added to the antibody as described in Example 1. The nanoreporter is the attached to the antibody by hybridization to the linker oligo at temperatures between 37 – 45°C. Antibody: reporter ratio was 1:1. The labeled antibody and the reporter were hybridized at 0.05 nM, 37°C, 1X SSPE, overnight.

[00177] A second direct approach is to first bind the antibodies to the target in solution and purify the complex as described in Example 1. After the purification the nanoreporter is hybridized to the antibody (after target binding) using the protocol described above. In this approach, problems of mismatch between hybridization and protein binding are avoided.

[00178] For both these approaches, strong binding affinities are also needed to remain bound during purifications and during imaging. Basically the Koff rate (antibody dissociation from the target protein) must be longer than the time for purification and reading of the assay.

[00179] Purification of the target protein-antibodies complex can be performed as described in Example 1. Elution can be accomplished by melting off G and F beads or via digestion.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1247

However, one skilled in the art will understand that the melting off the complex might require optimization of conditions to allow for antibodies to remain bound but affinity tags to release.

[00180] The protein-antibodies complex can be bound to a coverslip coated with streptavidin (Optichem®, Accelr8 Technology Corporation), stretched and image as described in Example 1.

[00181] Stretching of a reporter hybridized to an antibody was tested to be sure that the presence of the antibody does not cause non-specific binding or stickiness to the surface that inhibited the normal binding, stretching, and imaging process (data not shown).

Example 3. Detection of proteins using indirect measurement – Sandwich assay on a surface

[00182] A diagram of the protocol for this Example is depicted in Figure 6. In this Example the capture antibody is attached to a surface, e.g. a magnetic bead, and the second antibody is attached to a signal oligo. The antibodies can be prepared by any methods knows in the art including the methods described in Example 1. In this example, proteins and nanoreporters are largely separate, which eliminates concerns about protein stickiness. In this assay the local antibody concentrations on the surface can be high.

[00183] The target proteins are mixed with the capture antibody on the magnetic beads (2 hours to overnight, 1X PBS, and room temperature). The unbound protein sample is washed away. The labeled antibody signal oligo complex is added to the beads with blockers (1X PBS and room temperature). After a period of binding the excess labeled antibody signal oligo complex is washed away. The isolated target protein/antibody complex is then washed and the signal oligo is eluted and analyzed as described in Example 1.

Example 6. Detection of proteins using indirect measurement – Biotinylated signal oligos

[00184] A diagram of the protocol for this Example is depicted in Figure 7. In this Example the capture antibody is attached to a surface, e.g. a magnetic bead, and the second antibody is in solution or attached to a biotinylated signal oligo. This assay provides the advantages that it requires only two bead purifications. In addition, in this assay, like in Example 2, proteins and nanoreporters are largely separate, which eliminates concerns about protein stickiness. The target proteins are mixed with the capture antibody, labeled antibody signal oligo complex, and blockers as described in Example 3. The complex of target protein and antibodies are purified using the magnetic bead in the capture antibody as described in Example 3.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1248

[00185] The isolated target protein/antibody complex is then washed and the signal oligo is eluted as described in Example 1. The signal oligo can then be purified using Streptavidin-coupled Dynabeads® (Invitrogen) according to manufacturer's protocol. The signal oligo is then analyzed as described in Example 1.

Example 7. Detection of proteins using proximity ligation – Indirect Measurement

[00186] A diagram of the protocol for this Example is depicted in Figure 8. In this assay two physically close oligos are ligated. Probes containing the oligos are designed to bind pairwise to target proteins and to form a signal oligo by ligation when the probes are brought in proximity.

[00187] This approach has several benefits around sensitivity, minimization of cross-reactivity, and multiplexing. Proximity ligations have shown high sensitivity and have the effect of lowering the apparent Kd by essentially decreasing the off-rate.

[00188] The probes containing the oligos are prepared and purified as described in Gullberg et al, PNAS 101(22), p 8420-24 (2004). The target proteins are then mixed with the probes containing the oligos and the bridging oligos by incubating samples for one hour. The components required for probe ligation are then added as described in Gullberg et al. After five minutes ligation at room temperature the signal oligo is then released as via disulfide reduction, uracil excision, restriction digest, proteinase K, or any other suitable method know in the art. Additionally, the signal oligo can be released by the methods depicted in Figure 8B-8D

[00189] The signal oligo is the analyzed as described in Example 1.

[00190] Alternatively, the assay can be performed as depicted in Figure 9. In this approach, one of the oligos is attached to a nanoreporter.

[00191] This approach takes advantage of the decrease in the Koff via proximity ligation. A lower Koff means a lower Kd and the ability to work with lower concentrations of protein probe. This decrease in Kd makes it easier to work in concentrations required for reporters, and thus to contemplate direct detection approaches for multiplex analysis and lower reagent costs. This approach does not need a step for hybridization to reporters within the assay as some of the other methods proposed herein. Thus, it will be faster and have a shorter time to answer.

[00192] Using the approach described in Figure 8C, purification conditions are optimized to eliminate ligated oligos from non-ligated oligos after release from the antibodies. For example, this purification step can be performed using magnetic beads. Importantly, if the amount of

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1249

antibody used is higher than the amount of reporters used, then the resultant excess of unligated oligos may block the hybridization of the reporter to the oligo.

[00193] Antibody pairs were labeled with oligos using the methods described in Example 8. These oligos were designed to include a bridging oligo having an overlap of 9 bases and melting temperatures of 37° C in 1X PBS. The bridging oligo had 18 bases and a biotin tag for purification purposes. The ligated oligos have 18 bases of overlap with the biotinylated bridge and, thus, the ligated oligos are more stable than the unligated oligos that are bound to the antibodies. The biotinlyated oligos are isolated from the solution on magnetic beads coated with streptavidin. It was determined that only ligated oligos have sufficiently high melting temperatures to remain attached to the biotinylated oligos following a rinsing step using stringent buffer conditions.

[00194] Figure 13 demonstrates how various components are purified in a variety of buffer conditions. Components were present in solution at the concentrations used in the assay. The solutions were digested by protease prior to purification to release the oligos from the antibodies, as performed in the assay. As expected by melting temperature estimates, 0.03X SSPE provided the most efficient buffer. PROX05 represents a ligated product that is retained following this rinsing step.

Example 8. Detection of proteins using indirect measurements – Multiplexed assay in solution

[00195] This example is similar to Example 1, however, the detection is multiplexed and utilizes a different coupling chemistry between the antibodies and oligos.

Bioinformatics

[00196] Signal oligos were designed with minimal cross-reactivity at a range of temperatures, typically from about 4 to about 37°C,and 1X PBS. Unique overlaps between signal oligos and labeling oligos had melting temperatures of 51 to 56 in 1X PBS to allow for hybridizing the signal oligo to the labeling oligo at 37C. These overlaps of 15 to 17 bases had melting temperatures of 41 to 45°C in 0.1X SSPE to allow for elution after a magnetic bead purification.

Coupling of oligo to antibodies

[00197] Oligos were coupled to antibodies using aldehyde-hydrazine chemistry. All antibodies and target were purchased from R&D systems, Inc. (Minneapolis, MN). Each antibody A (see, Table 1) was desalted using a size exclusion spin column (0.5 ml Zeba Spin Column, Fisher Scientific, Pittsburgh, PA). Succinimidyl 6-hydrazinonicotinate acetone hydrazone (Solulink,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1250

San Diego, CA) was reacted to each antibody A. Each antibody was again purified using a size exclusion spin column.

[00198] Table 1. Antibodies

		Part numbers	
Protein	**Target**	**Antibody A**	**Antibody B**
TNF alpha	210-TA-010	MAB610	BAF210
IL1 alpha	200-LA-002	MAB200	BAF200
IL6 alpha	206-TA-010	MAB206	BAF206
VEGF	293-VE-010	MAB293	BAF293

[00199] The amine oligos were desalted using a membrane spin column (5K MWCO VivaSpin, Fisher Scientific, Pittsburgh, PA). Twenty molar equivalents succinimidyl-4-formyl benzoate (Solulink, San Diego, CA) were reacted with each oligo. The oligos were again purified using membrane spin columns.

[00200] Each corresponding modified oligo was reacted with the corresponding modified antibody at a molar ration of 3 to 1. This was purified on a spin column (2 ml Zeba Spin Column, Fisher Scientific, Pittsburgh, PA). Table 2 shows that between 1-2 oligos were attached to each antibody at the end of the counting process (Oligo:Ab).

[00201] Table 2. Molar substitution ratios showing quality control and quantification capabilities of antibody – oligo coupling process.

Antibody	**MSR (molar substitution ratio)**		
Protein target	4FB:oligo	HyNic:Ab	Oligo:Ab
TNFalpha	1.46	7.3	1.48
IL1alpha	2.31	11	1.53
VEGF	3.58	7.0	2.07
IL6	1.34	7.4	1.37

Hybridization of signal oligo to detection antibody

[00202] Each unique signal oligo was preannealed separately to the unique oligo linked to each antibody A by adding oligo linked antibody A and corresponding signal oligo in a ratio of 4:1, signal oligo:Antibody ratio. Other ratios are contemplated.

Formation of target and antibodies complex

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1251

[00203] A 2X master mix was created containing: each antibody A annealed to signal oligo and biotinylated antibody B (4 pairs) in a single tube at approximately $1x10^{-15}$ to $1x10^{-8}$ M, with blockers (salmon sperm). The target proteins were added to the desired dilution ($<1x10^{-8}$M) to aliquots of this master mix. Antibodies were at 10X concentration over an estimated Kd of 10^{-15} to 10^{-8} M. The mixture was incubated.

[00204] The complex of target protein and antibodies was purified using Streptavidin-coupled Dynabeads® (Invitrogen) according to protocol.

Elution of Signal Oligo

[00205] The isolated target protein/antibody complex was washed and the signal oligo was eluted at greater than 45°C using 0.1X SSPE for 10 to 15 minutes. Shorter and longer periods are contemplated

Detection of Signal Oligo:

[00206] Detection of the signal oligo in each sample was carried out using a dual nanoreporter system having both labeled nanoreporter probes and unlabeled nanoreporter probes. The signal oligos from each sample were hybridized with final concentrations of the hybridization reagents as follows: unlabeled, biotinylated probe labeled reporter probe, 5X SSPE (pH 7.5), 5X Denhardt's reagent (Sigma), sheared salmon sperm DNA (Sigma), and detergent. Reagents were mixed and incubated in a thermocycler block with a heated lid for 16 hours.

Post-hybridization purification

[00207] To remove unhybridized reporters, reactions were purified over magnetic beads (InvitrogenTM) coupled to oligonucleotides complementary to the 3'-repeat sequence contained on every biotinylated probe. Reactions were first diluted SSPE in 0.1% detergent mixture/TE and allowed to bind to beads at greater than 20C with continuous rotation. The beads were washed three times in SSPE and detergent and the hybridized complexes eluted in of 0.1X SSPE/0.1% /detergent mixture for 15 minutes at 45°C. After elution, samples were purified a second time to remove excess biotinylated probes by binding to magnetic beads coupled to oligonucleotides complementary to the 5'-repeat sequence contained on every reporter probe. The elutions from the anti-3'-repeat beads were brought to a final concentration of 1X SSPE and bound for 15 minutes at 22.5°C with rotation. Beads were washed as above and eluted in of 0.1X SSPE/0.1% /detergent mixture at greater than 40°C. The doubly-purified samples were then prepared for capture as described below.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1252

NanoString reporter capture, stretching, and imaging

[00208] A solution of a custom-formulation of Tetraspeck fluorescent microspheres (InvitrogenTM) was added to each sample. Samples were loaded into a NanoString fluidic device processed and imaged.

Results

[00209] Results are shown in Figure 10. The results of this experiment showed that 4 proteins were simultaneously detected by the assay described herein (Figure 10). The proteins detected were TNFα, IL1α , IL6, and VEGF. This experiment showed sensitivities of about $1x10^{-12}$ M. Figure 11 shows the same data plotted versus fluid sample. A subsequent experiment (Figure 12) showed the limit of detection of two of these proteins was 26 and 38 pg/ml ($1.4x10^{-12}$ and $1.9x10^{-12}$M for IL1α and IL6, respectively).

[00210] Improvements in background, *i.e.* reduction of background detection or the improvement of background to target detection ratios, allow for increases in sensitivities of approximately 2 orders of magnitude, thus sensitivity would reach $1x10^{-14}$ M or significantly <1 pg/ml.

Example 9. Anti-Streptavidin Probe Reporter

[00211] An anti-streptavidin antibody (Affinity Bioreagents, Rockford, IL) was labeled with an oligonucleotide (oligo), as described in Example 1.

[00212] In a particular example of this embodiment, an antibody-labeled oligo was hybridized to a reporter at concentration of 0.05 nM, at a temperature of 45°C, in 1X SSPE buffer, overnight. There was a 25-base overlap between the reporter and oligo that was bound to the antibody. This overlap may be shortened, if desired. In certain embodiments, the overlap is optionally 1, 5, 10, 15, 20, 25 bases or any length in between. A shorter overlap of bases between the reporter and oligo on the antibody allows for increased efficiency of hybridization of the antibody-oligo to the reporter at temperatures that produce antibody stability. A reporter with an anti-streptavidin antibody probe was introduced into the flow chamber of the cartridge (at a concentration of 0.025 nM in 0.25X SSPE buffer) and a first end of the reporter was allowed to bind to the streptavidin surface for 10 minutes. The chamber was then washed with TAE buffer. The reporters were stretched and subsequently immobilized on the surface by first using an electric field of 200 volts (V)/centimeter (cm) and, second, introducing biotinlyated oligos to

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1253

attach a second end of the reporter to the surface. The sample was washed again with TAE, and SlowFade™ was introduced to stabilize the dyes. This sample was then imaged.

[00213] Figure 14 shows that only the reporter with the anti-streptavidin probe (S16) was detected.

OTHER EMBODIMENTS

[00214] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1254

CLAIMS

WHAT IS CLAIMED IS:

1. A method for determining the concentration of at least one protein in a sample comprising the steps:

(a) providing:

(i) at least one protein,

(ii) a first protein probe specific for a first region of said at least one protein, wherein said first protein probe comprises a capture region,

(iii) a second protein probe specific for a second region of said at least one protein, wherein said second protein probe comprises a nanoreporter comprising a plurality of different detectable labels, and

(iv) a matrix having attached thereto a moiety which is capable of binding to said capture region in said first protein probe

(b) forming at least a complex comprising said at least one protein, said first protein probe, said second protein probe and said moiety, wherein said at least one protein is bound to said first and second protein probes, and wherein said moiety is bound to said capture probe in said first protein probe; and

(c) individually detecting said complex or at least part of said complex by a method comprising individually counting the presence of one or more molecules of said nanoreporter wherein the presence of said one or more molecules is indicative of the concentration of said protein in the sample.

2. The method of claim 1 wherein said individually detecting further comprises detecting a digital signal.

3. The method of claim 1 further comprising determining the concentration of a plurality of target proteins by a method comprising forming a plurality of complexes, each complex comprising (i) at least one target protein (ii) a first protein probe specific for a first region of said at least one protein, wherein said first protein probe comprises a capture region (iii) a second protein probe specific for a second region of said at least one protein, wherein said second

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1255

protein probe comprises a nanoreporter comprising a plurality of different detectable labels and (iv) a moiety attached to a matrix, wherein said moiety is capable of binding to said capture region in said first protein probe, wherein each second protein probe comprises a different nanoreporter region.

4. The method of claim 3, wherein each nanoreporter in said plurality of complexes has a detectable signal that distinguishes it from other nanoreporters in said population.

5. The method of claim 3 wherein the dissociation constant of said first and said second protein probes is about $1.00x10^{-15}$ to about $1.00x10^{-08}$.

6. The method of claim 3 wherein the concentration of two or more target proteins is determined.

7. The method of claim 6 wherein the concentration of 3, 4, 5, 10, 20, 30, 50, 100, 200, 300, 500, 600, 700, 800, 900, 1000 or more than 1000 different target proteins is determined.

8. The method of claim 3 wherein the concentration up to 2000 different target proteins is determined.

9. The method of claim 3 wherein the concentration up to 980 different target proteins is determined

10. The method of claim 1 wherein the matrix is selected from the group consisting of a bead and an array.

11. The method of claim 3 wherein the matrix is a bead and wherein each moiety in each complex of said plurality of complex is attached to a different bead.

12. The method of claim 3 wherein the matrix is a surface and wherein each moiety in each complex of said plurality of complex is attached to a different location of said surface.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1256

13. The method of claim 1 wherein said first protein probe and said second protein probe are independently selected from the group consisting of antibody, peptide, aptamer and peptoid.

14. The method of claim 1 wherein said nanoreporter comprises a single-stranded nucleic acid backbone, said backbone comprising a plurality of label attachment regions covalently attached together in a linear combination, wherein each label attachment region is hybridized to a complementary polynucleotide sequence having attached thereto the detectable label.

15. The method of claim 1 wherein said nanoreporter is attached to said second probe through hybridization to a linker oligo.

16. The method of claim 15 wherein said nanoreporter is hybridized to said linker oligo at a temperature of about 32 to about 40 degrees Celsius.

17. The method of claim 15 wherein said nanoreporter is hybridized to said linker oligo at a temperature of about 37 to 45 degrees Celsius.

18. The method of claim 15 wherein said nanoreporter is hybridized to said linker oligo at a temperature of about 37 degrees Celsius.

19. The method of claim 15 wherein said nanoreporter comprises a portion that is complementary to the linker oligo.

20. The method of claim 19 wherein said complementary region is about 15 to about 20 bases.

21. A method for determining the concentration of at least one protein in a sample comprising the steps:

(a) providing:

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1257

(i) at least one protein,

(ii) a first protein probe specific for a first region of said at least one protein, wherein said first protein probe is attached to a first capture region or a first matrix

(iii) a second protein probe specific for a second region of said at least one protein, wherein said second protein probe comprises a signal oligo; and

(iv) when said first probe is attached to a first capture region: a second matrix having attached thereto a moiety which is capable of binding to said capture region in said first protein probe.

(b) forming at least a first complex comprising said at least one protein, said first protein probe, and said second protein probe, wherein said at least one protein is bound to said first and second protein probes, and wherein when said first probe is attached to a first capture region said capture probe is bound to said moiety in said second matrix;

(c) releasing said signal oligo from said first complex;

(d) forming a second complex comprising (1) at least said signal oligo and (2) at least one oligo probe comprising a signal oligo-specific region and a region comprising a nanoreporter wherein said nanoreporter comprises a plurality of different detectable labels;

(e) individually detecting said second complex or at least part of said second complex by a method comprising individually counting the presence of one or more molecules of said nanoreporter, wherein the presence of said second one or more molecules is indicative of the concentration of said protein in said sample.

22. The method of claim 21 wherein said signal oligo is attached to a second capture region.

23. The method of claim 22 wherein said releasing of said signal oligo further comprises capturing directly or indirectly said signal molecule into a third matrix.

24. The method of claim 23 wherein said nanoreporter further comprising a constant region, wherein the constant region comprises a plurality of repeat nucleotide sequences.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1258

25. The method of claim 24 further comprising binding said constant region to a second moiety in said third matrix, wherein said second moiety is capable of binding said constant region.

26. The method of claim 21 wherein said individually detecting further comprises detecting a digital signal.

27. The method of claim 21 wherein the first matrix is selected from the group consisting of a bead and an array.

28. The method of claim 21 wherein the second matrix is selected from the group consisting of a bead and an array.

29. The method of claim 21 further comprising determining the concentration of a plurality of target proteins by a method comprising forming a plurality of complexes, each complex comprising (i) at least one target protein (ii) a first protein probe specific for a first region of said at least one protein, wherein said first protein probe is attached to a capture region or a first matrix (iii) a second protein probe specific for a second region of said at least one protein, wherein said second protein probe comprises a signal molecule, wherein when said first probe is attached to a first capture region said capture probe is bound to said moiety in said second matrix, and wherein each second protein probe in each said plurality of complexes comprises a different signal oligo.

30. The method of claim 29 wherein the first matrix is a bead and the bead comprises a plurality of identical first protein probes.

31. The method of claim 29 wherein the concentration of two or more target proteins is determined.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1259

32. The method of claim 29 wherein the concentration of 2, 3, 4, 5, 10, 20, 30, 50, 100, 200, 300, 500, 600, 700, 800, 900, 1000 or more than 1000 different target proteins is determined.

33. The method of claim 29 wherein the concentration up to 2000 different target proteins is determined.

34. The method of claim 29 wherein the concentration up to 980 different target proteins is determined.

35. The method of claim 21 wherein said first protein probe and said second protein probe are independently selected from the group consisting of antibody, peptide, aptamer and peptoid.

36. The method of claim 21 wherein said nanoreporter comprises a single-stranded nucleic acid backbone, said backbone comprising a plurality of label attachment regions covalently attached together in a linear combination, wherein each label attachment region is hybridized to a complementary polynucleotide sequence having attached thereto the detectable label.

37. A method for determining the concentration of at least one protein in a sample comprising the steps:

(a) providing:

(i) at least one protein,

(ii) a first protein probe specific for a first region of said at least one protein, wherein said first protein probe is attached to a first oligo

(iii) a second protein probe specific for a second region of said at least one protein, wherein said second protein probe is attached to a second oligo;

(b) forming a first complex comprising said at least one protein, said first protein probe and said second protein probe, wherein said at least one protein is bound to said first and second protein probes,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1260

(c) ligating said first and said second oligo to form a signal oligo

(d) forming a second complex comprising (1) said first signal oligo and (2) a at least one oligo probe comprising a signal oligo-specific region and a region comprising a nanoreporter wherein said nanoreporter comprises a plurality of different detectable labels

(e) individually detecting said second complex or at least part of said second complex by a method comprising individually counting the presence of one or more molecules of said nanoreporter, wherein the presence of said one or more molecules is indicative of the concentration of said protein in said sample.

38. The method of claim 37 further comprising releasing said signal oligo from said first complex.

39. The method of claim 38 wherein said signal oligo comprises a capture region.

40. The method of claim 39 wherein said releasing of said signal oligo further comprises capturing directly or indirectly said signal molecule into a matrix.

41. The method of claim 37 wherein said individually detecting further comprises detecting a digital signal.

42. The method of claim 37 further comprising determining the concentration of a plurality of target proteins by a method comprising forming a plurality of complexes, each complex comprising (i) at least one target protein (ii) a first protein probe specific for a first region of said at least one protein, wherein said first protein probe is attached to a first oligo (iii) a second protein probe specific for a second region of said at least one protein, wherein said second protein probe is attached to a second oligo, wherein the ligation of said first oligo and said second oligo form a signal oligo, and wherein each complex in said plurality of complexes comprises a different signal oligo.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1261

43. The method of claim 42 wherein the concentration of two or more target proteins is determined.

44. The method of claim 43 wherein the concentration of 2, 3, 4, 5, 10, 20, 30, 50, 100, 200, 300, 500, 600, 700, 800, 900, 1000 or more than 1000 different target proteins is determined.

45. The method of claim 41 wherein the concentration of at least 972 different target proteins is determined.

46. The method of claim 37 wherein said first protein probe and said second protein probe are independently selected from the group consisting of antibody, peptide, aptamer and peptoid.

47. The method of claim 37 wherein said nanoreporter comprises a single-stranded nucleic acid backbone, said backbone comprising a plurality of label attachment regions covalently attached together in a linear combination, wherein each label attachment region is hybridized to a complementary polynucleotide sequence having attached thereto the detectable label.

48. A population of uniquely labeled protein probes, wherein each probe comprises

i) a target-specific region; and

ii) a region comprising a nanoreporter comprising a plurality of different detectable molecules;

wherein said nanoreporter in each protein probe has a detectable signal that distinguishes it from other nanoreporters in said population.

49. The population of claim 48 wherein target-specific region is selected from the group consisting of antibody, peptide, aptamer and peptoid.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1262

50. The population of claim 48 wherein said nanoreporter comprises a single-stranded nucleic acid backbone, said backbone comprising a plurality of label attachment regions covalently attached together in a linear combination, wherein each label attachment region is hybridized to a complementary polynucleotide sequence having attached thereto the detectable label.

51. The population of claim 48 wherein the dissociation constant of the target specific region is about $1.00x10^{-15}$ to about $1.00x10^{-08}$.

52. The population of claim 48 wherein said nanoreporter is attached to said protein probe through hybridization to a linker oligo.

53. The population of claim 52 wherein said nanoreporter is hybridized to said linker oligo at a temperature of about 32 to about 40 degrees Celsius.

54. The population of claim 53 wherein said nanoreporter is hybridized to said linker oligo at a temperature of about 37 degrees Celsius.

55. The population of claim 52 wherein said nanoreporter comprises a portion that is complementary to the linker oligo.

56. The method of claim 55 wherein said complementary region is about 15 to about 20 bases.

57. The population of claim 49 wherein said nanoreporter is attached to said antibody through hybridization to a linker oligo and wherein said nanoreporter is hybridized to said linker oligo at a temperature of about 37 – 45 degrees Celsius.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1263

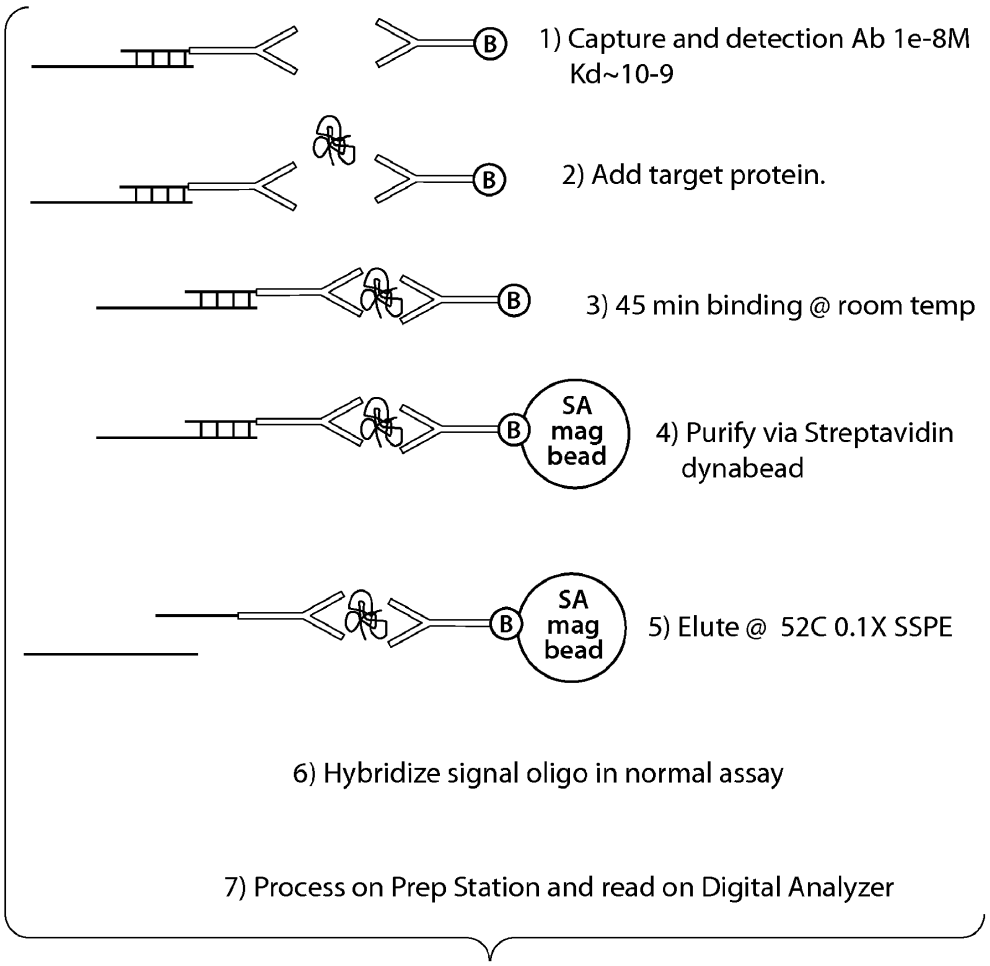


Fig. 1

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1264

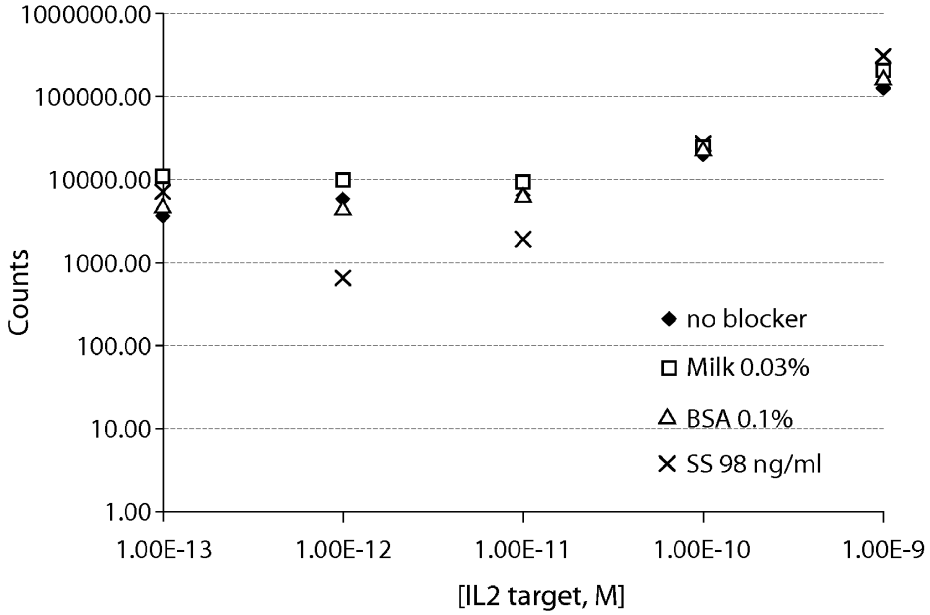


Fig. 2

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1265

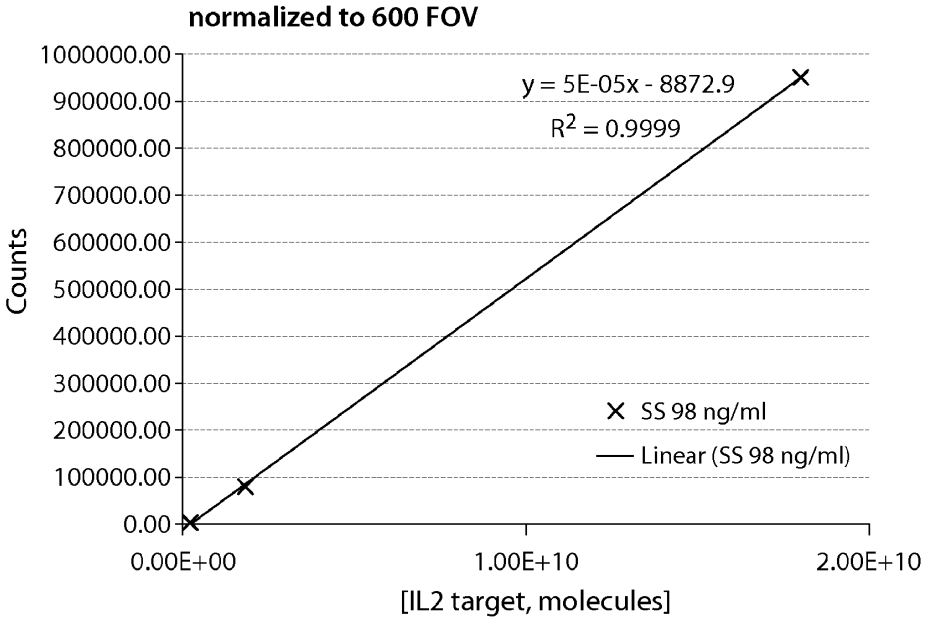


Fig. 3

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1266

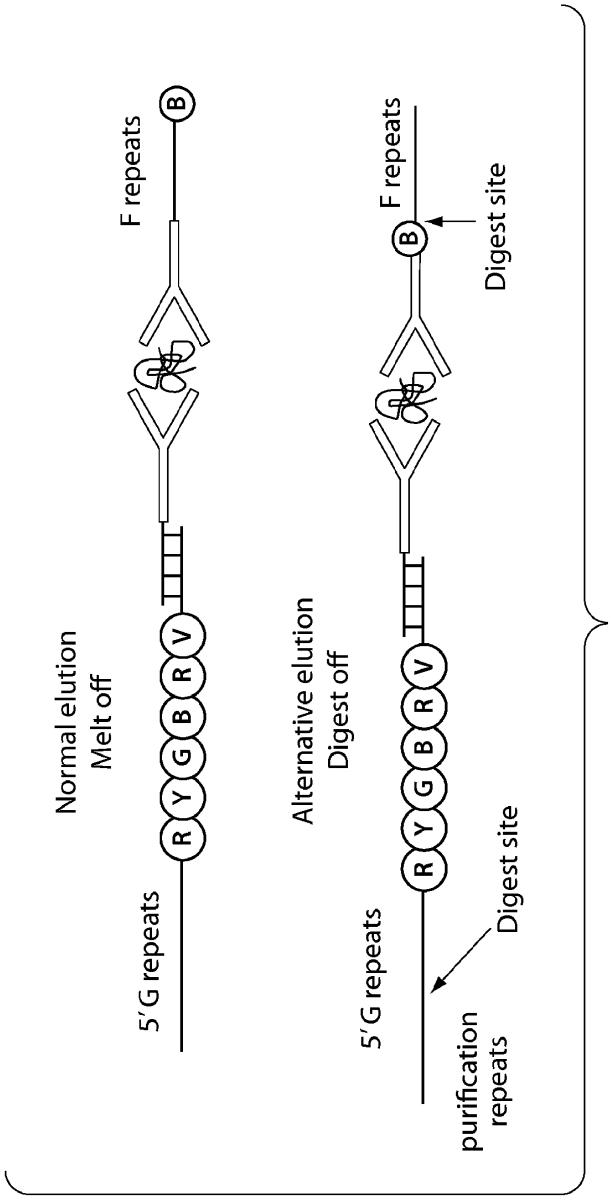


Fig. 4

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1267

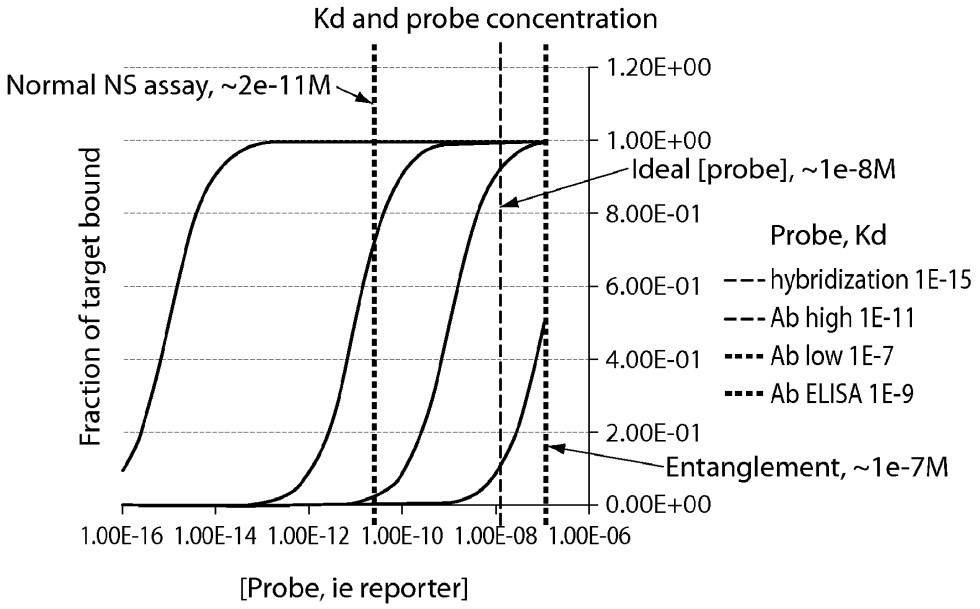


Fig. 5

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1268

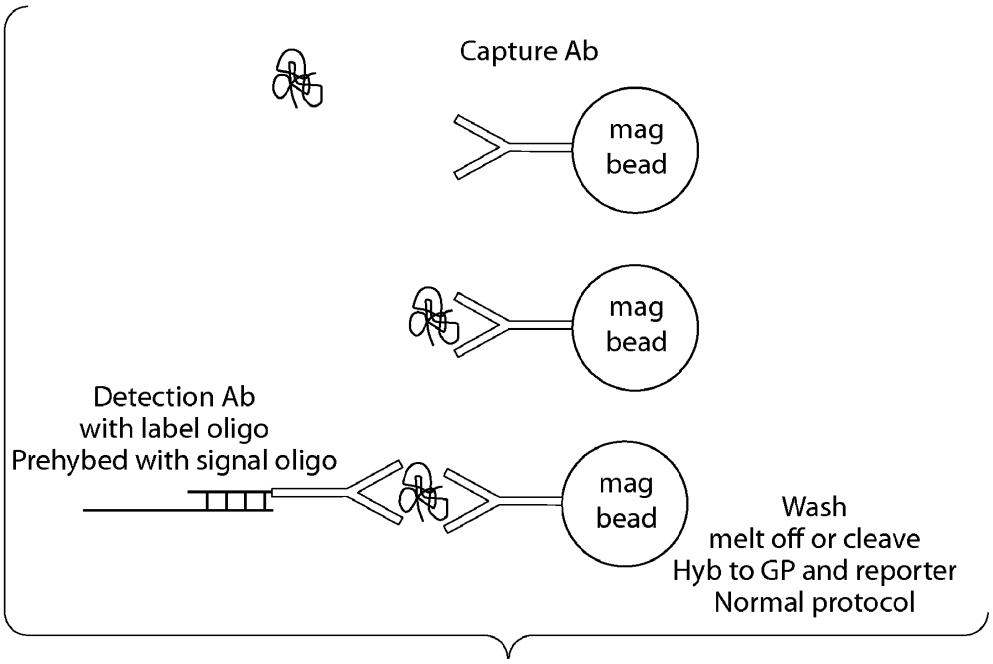


Fig. 6

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1269

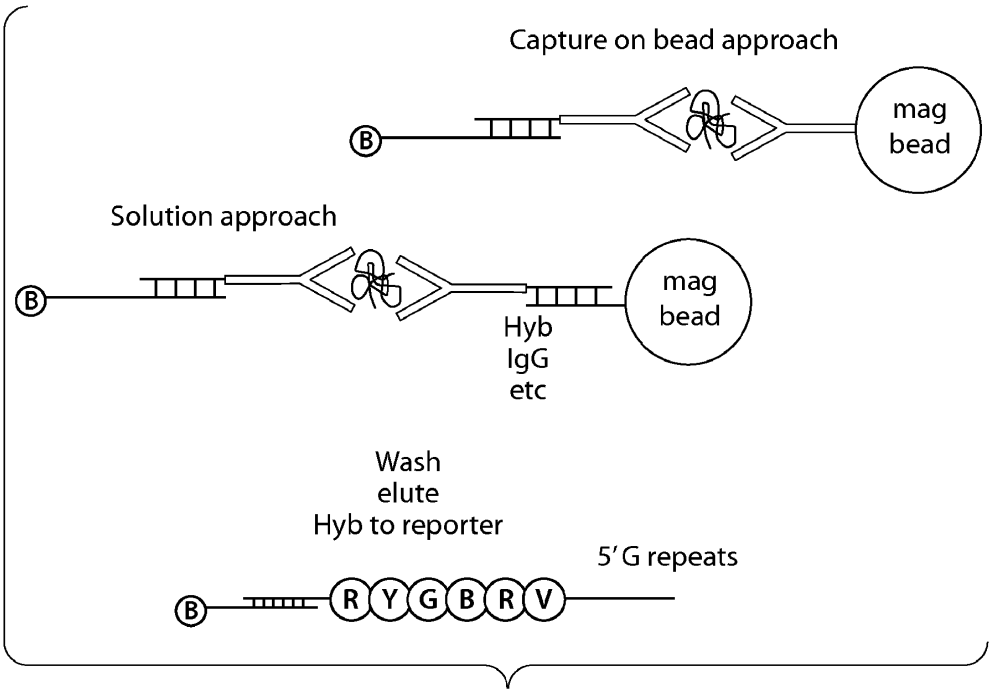


Fig. 7A

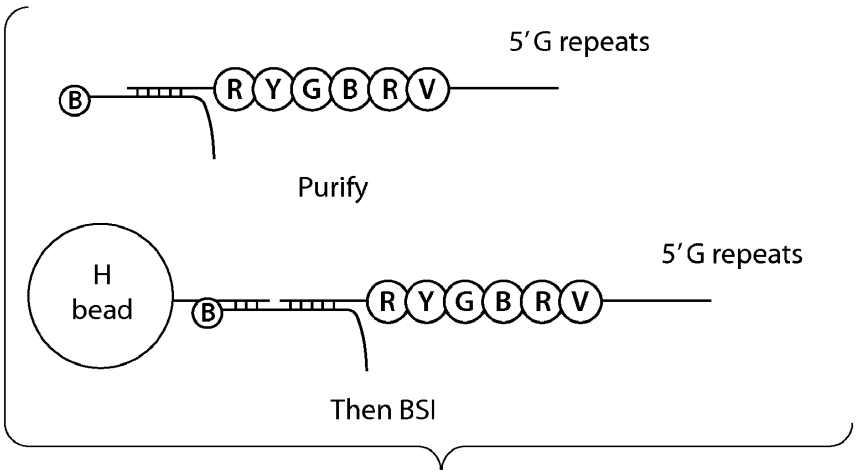


Fig. 7B

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1270

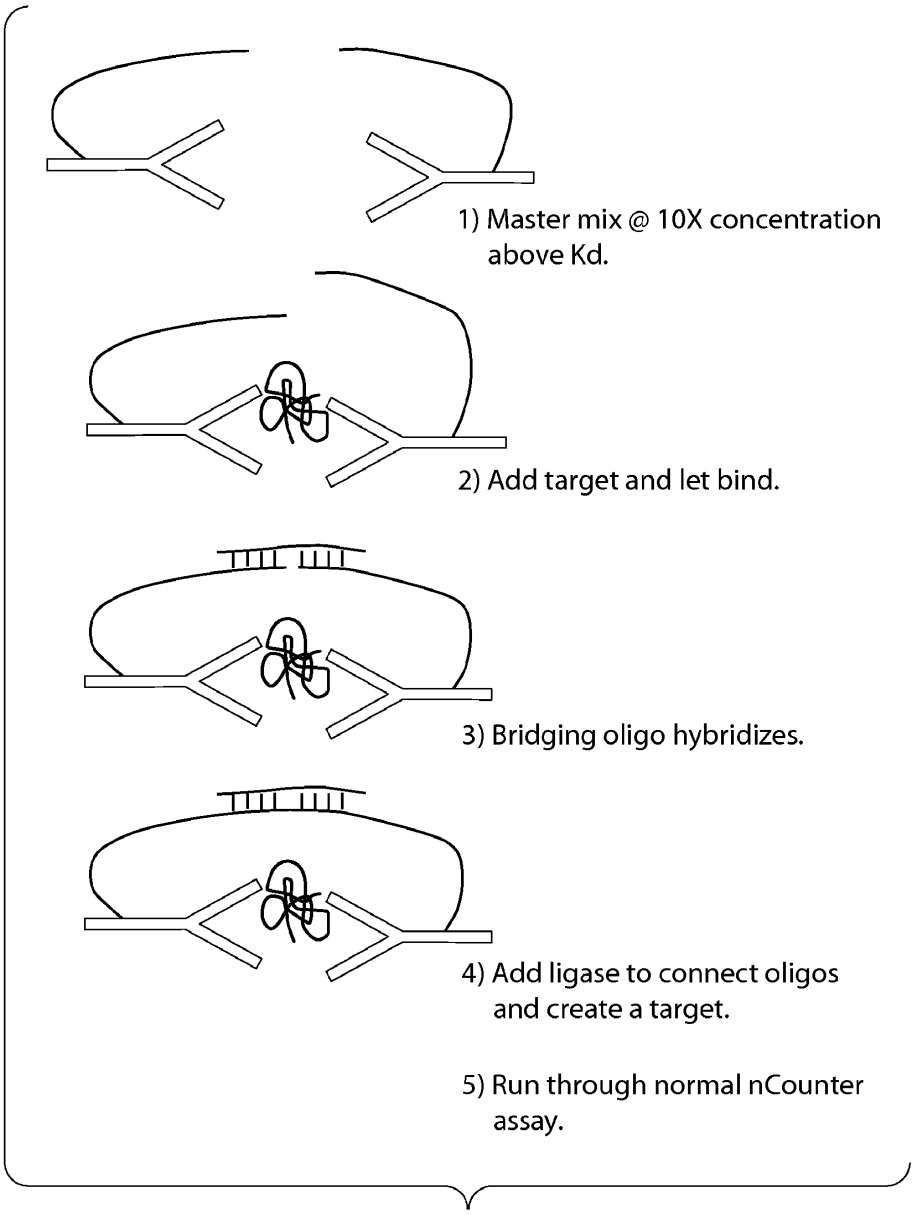


Fig. 8A

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1271

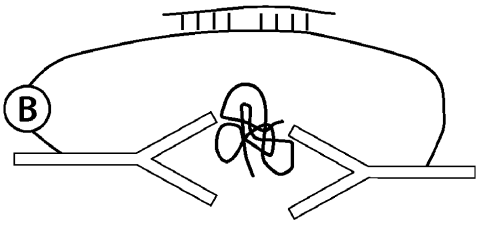


Fig. 8B

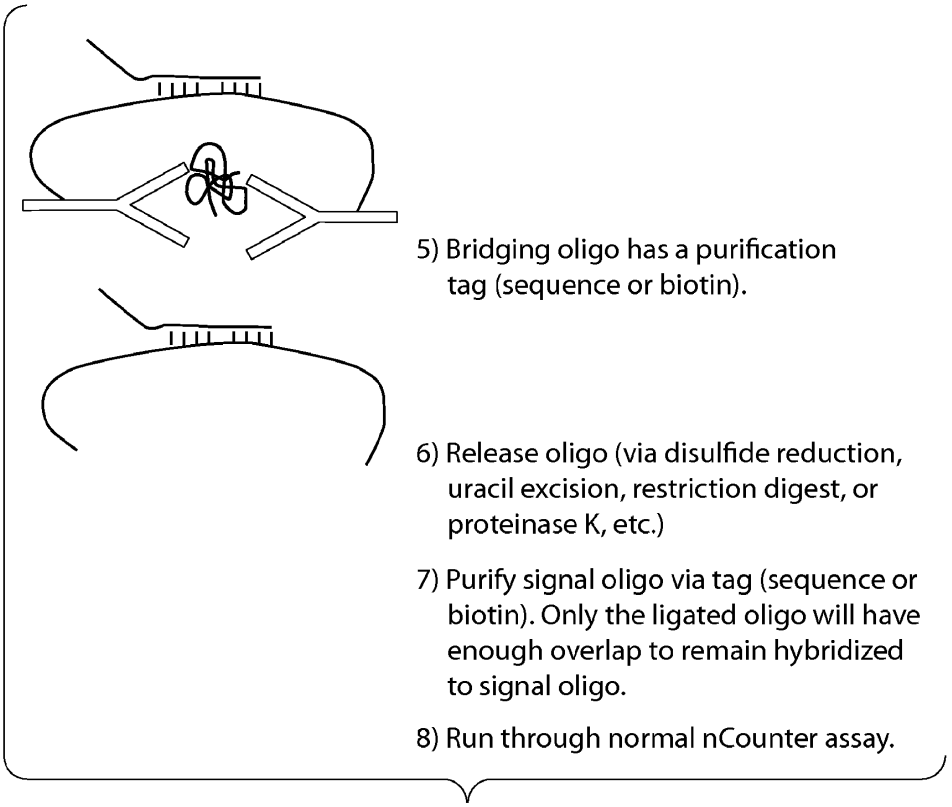


Fig. 8C

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1272

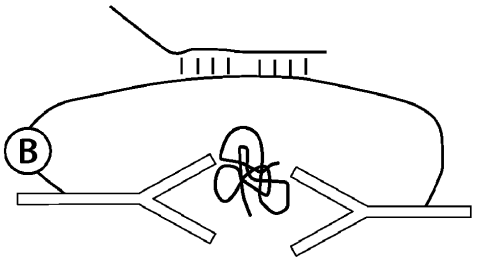


Fig. 8D

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1273

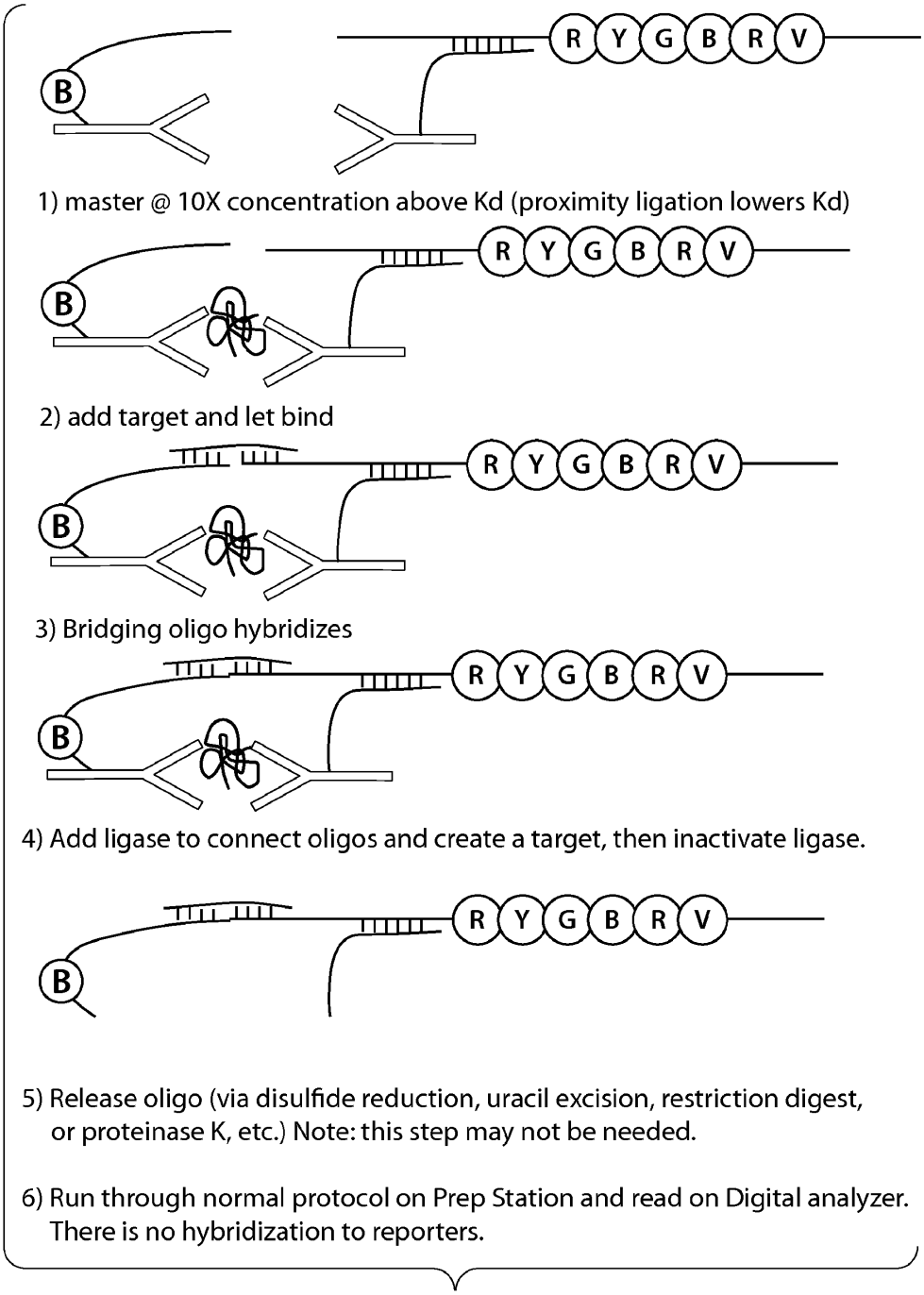


Fig. 9

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1274

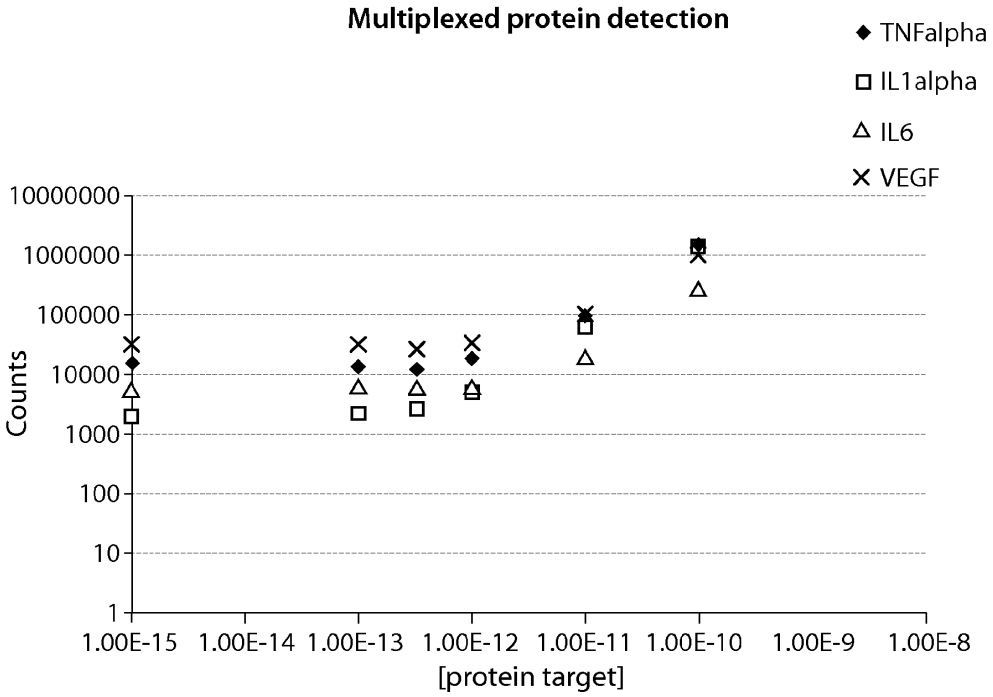


Fig. 10

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1275

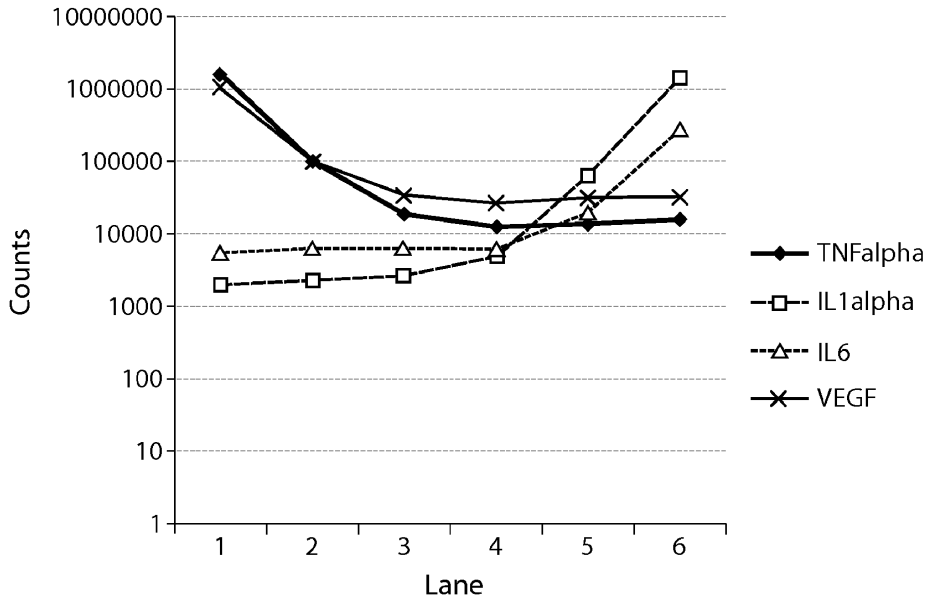


Fig. 11

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1276

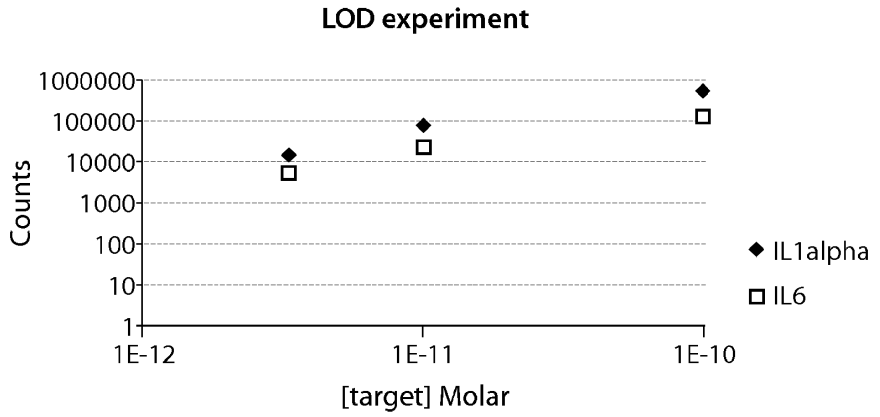


Fig. 12

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1277

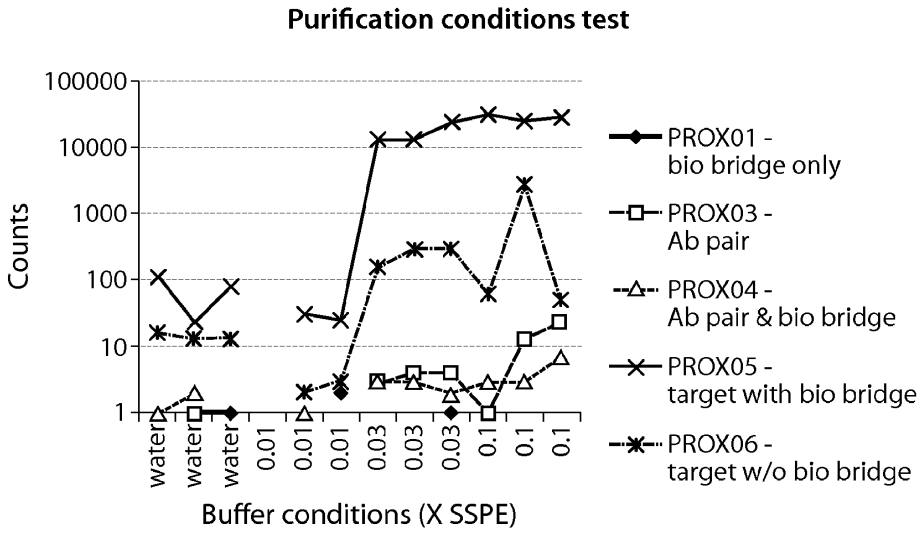


Fig. 13

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1278

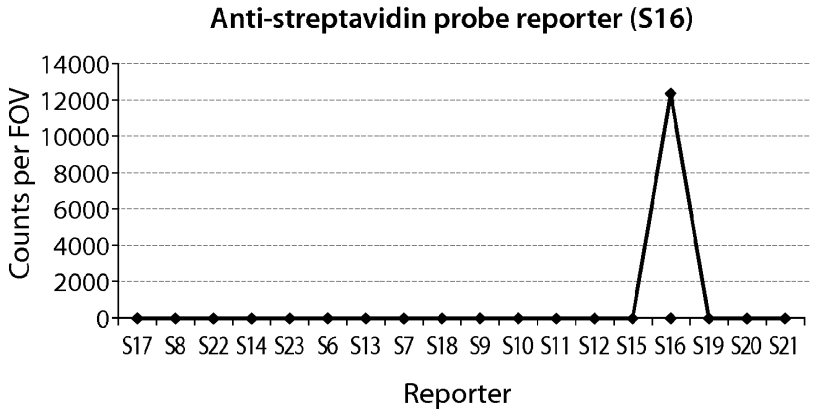


Fig. 14

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1279

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 April 2011 (21.04.2011)

PCT

(10) International Publication Number
WO 2011/047087 A3

(51) International Patent Classification:
G01N 33/68 (2006.01) *G01N 33/58* (2006.01)
G01N 33/53 (2006.01) *C12Q 1/68* (2006.01)

(21) International Application Number:
PCT/US2010/052556

(22) International Filing Date:
13 October 2010 (13.10.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/251,192 13 October 2009 (13.10.2009) US
61/325,224 16 April 2010 (16.04.2010) US
61/326,787 22 April 2010 (22.04.2010) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:
US 61/325,224 (CIP)
Filed on 16 April 2010 (16.04.2010)
US 61/251,192 (CIP)
Filed on 13 October 2009 (13.10.2009)
US 61/326,787 (CIP)
Filed on 22 April 2010 (22.04.2010)

(71) Applicant *(for all designated States except US)*: **NANOSTRING TECHNOLOGIES, INC.** [US/US]; 530 Fairview Avenue North, Suite 2000, Seattle, WA 98109 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **DUNAWAY, Dwayne, L.** [US/US]; 2751 N.E. 94th Street, Seattle, WA 98115 (US).

(74) Agents: **ELRIFI, Ivor, R.** et al.; Mintz Levin Cohn Ferris Glovsky And Popeo, P.C., One Financial Center, Boston, MA 02111 (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, 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, 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, PE, PG, PH, PL, PT, RO, RS, RU, 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,

[Continued on next page]

(54) Title: PROTEIN DETECTION VIA NANOREPORTERS

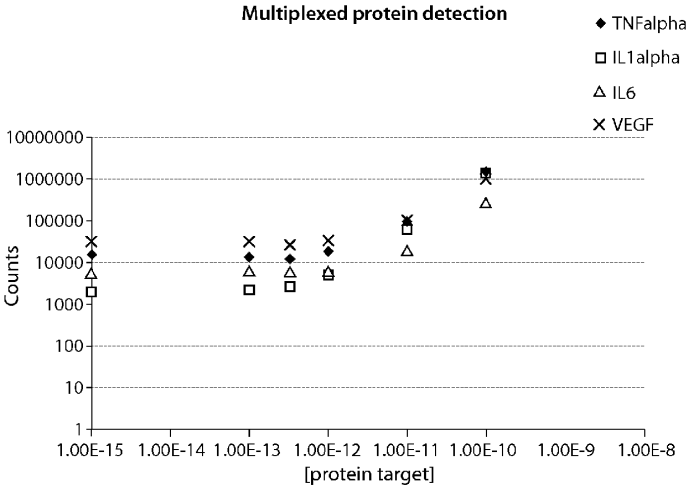


Fig. 10

(57) Abstract: The invention provides methods, compositions, kits and devices for the detection of proteins. In some embodiments, the invention allows for multiplexed protein detection.

WO 2011/047087 A3

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1280

GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

(88) Date of publication of the international search report:
18 August 2011

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2010/052556

A. CLASSIFICATION OF SUBJECT MATTER

G01N 33/68(2006.01)i, G01N 33/53(2006.01)i, G01N 33/58(2006.01)i, C12Q 1/68(2006.01)i

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)
G01N 33/68; C12Q 1/68

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Korean utility models and applications for utility models
Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PubMed

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y A	WO 2008/124847 A2 (GEISS, G. K. et al.) 16 October 2008. See page 1, lines 5-12.	48,49 1-13 14-47,50-57
Y A	BLANK, K. et al. Double-chip protein arrays: force-based multiplex sandwich immunoassays with increased specificity. Analytical and Bioanalytical Chemistry. 21 April 2004, Vol. 379, No.7-8, pages 974-981. See results; figures 1, 3.	1-13 14-57
A A	KANG, Y. et al. Electrochemical detection of thrombin by sandwich approach using antibody and aptamer. Bioelectrochemistry. Vol. 73, Issue 1, June 2008, pages 76-81. See figure 1. SASAJIMA, Y. et al. Detection of protein tyrosine phosphorylation by open sandwich fluoroimmunoassay. Biotechnology Progress. July-August 2006. Vol. 22, No. 4, pages 968-973. See whole documents.	1-57 1-57

☒ 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 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 JUNE 2011 (29.06.2011)	**30 JUNE 2011 (30.06.2011)**
Name and mailing address of the ISA/KR Korean Intellectual Property Office Government Complex-Daejeon, 189 Cheongsa-ro, Seo-gu, Daejeon 302-701, Republic of Korea Facsimile No. 82-42-472-7140	Authorized officer Lee Hyun Ji Telephone No. 82-42-481-8745

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2010/052556

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	HAN, X. X. et al. Protein-mediated sandwich strategy for surface-enhanced Raman scattering: application to versatile protein detection. Analytical Chemistry. May 2009, Vol. 81, No. 9, pages 3350-3355. See abstract.	1-57

Form PCT/ISA/210 (continuation of second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/US2010/052556

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2008-124847 A2	16.10.2008	AU 2008-237018 A1	16.10.2008
		CA 2687292 A1	16.10.2008
		EP 2155899 A2	24.02.2010
		JP 2010-523159 A	15.07.2010
		JP 2010-523159 T	15.07.2010
		US 2010-112710 A1	06.05.2010
		WO 2008-124847 A2	16.10.2008
		WO 2008-124847 A3	16.10.2008
		WO 2008-124847 A3	19.02.2009

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1284

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 August 2011 (18.08.2011)

PCT

(10) International Publication Number
WO 2011/100541 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01) *C12N 15/113* (2010.01)

(21) International Application Number:
PCT/US2011/024519

(22) International Filing Date:
11 February 2011 (11.02.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/303,517 11 February 2010 (11.02.2010) US

(71) Applicant *(for all designated States except US)*: **NANOSTRING TECHNOLOGIES, INC.** [US/US]; 530 Fairview Avenue North, Suite 2000, Seattle, WA 98109 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **WEBSTER, Philippa, Jane** [US/US]; 4317 Densmore Ave. N, Seattle, WA 98103 (US). **DENNIS, Lucas, Michael** [US/US]; 4542 Orcas Street, Seattle, WA 98118 (US). **BOYKIN, Richard, Kemble** [US/US]; 16741 10th Ave NW, Shoreline, WA 98177 (US).

(74) Agents: **ELRIFI, Ivor, R.** et al.; Mintz, Levin, Cohn, Ferris, Glovsky & Popeo, P.C., One Financial Center, Boston, MA 02111 (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, 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, 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, PE, PG, PH, PL, PT, RO, RS, RU, 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.

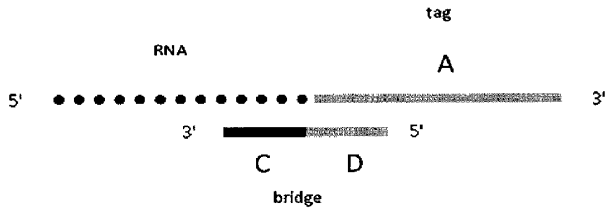
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, 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, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

(54) **Title**: COMPOSITIONS AND METHODS FOR THE DETECTION OF SMALL RNAS

FIGURE 2



(57) **Abstract**: The invention provides compositions and methods for the detection of small RNA molecules in a multiplexed reaction. The assays and kits described herein are applicable for the identification, diagnosing, and monitoring of disorders including, but not limited to cancer, developmental and degenerative disease, neurological disorders, and stem cell disorders.

WO 2011/100541 A2

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1285

COMPOSITIONS AND METHODS FOR THE DETECTION OF SMALL RNAS

RELATED APPLICATIONS

[01] This application claims priority from U.S. Provisional Patent Application 61/303,517 filed on February 11, 2010 and incorporated, herein, in its entirety.

FIELD OF THE INVENTION

[02] This invention relates generally to the field of molecular biology, and specifically, to the fields of detection, identification, and quantification of target nucleic acid molecules in mixtures.

BACKGROUND OF THE INVENTION

[03] Although all cells in the human body contain the same genetic material, the same genes are not active in all of those cells. Alterations in gene expression patterns can have profound effects on biological functions. These variations in gene expression are at the core of altered physiologic and pathologic processes. Therefore, identifying and quantifying the expression of genes in normal cells compared to diseased cells can aid the discovery of new drug and diagnostic targets.

[04] Nucleic acids can be detected and quantified based on their specific polynucleotide sequences. The basic principle underlying existing methods of detection and quantification is the hybridization of a labeled complementary probe sequence to a target sequence of interest in a sample. The formation of a duplex indicates the presence of the target sequence in the sample.

[05] This technique, called molecular hybridization, has been a useful tool for identifying and analyzing specific nucleic acid sequences in complex mixtures. This technique has been used in diagnostics, for example, to detect nucleic acid sequences of various microbes in biological samples. In addition, hybridization techniques have been used to map genetic differences or polymorphisms between individuals. Furthermore, these techniques have been used to monitor changes in gene expression in different populations of cells or in cells treated with different agents.

[06] The identification of small RNA molecules, microRNA (miRNA) molecules, with hybridization techniques presents several unique challenges. Hybridization of a detection probe to a short length RNA molecule of a typical miRNA target occurs at low melting temperatures and prevents concurrent binding by multiple detection moeities. Low melting temperature

hybridizations are unfavorable for specific binding of multiple probes to target miRNA sequences that may differ by only a single nucleic acid. Furthermore, miRNA sequences demonstrate great diversity despite having constant or conserved lengths, which generates a large variety of hybridization melting temperatures.

[07] Thus, there exists a need for accurate and sensitive detection, identification and quantification of target nucleic acid molecules in complex mixtures. Particularly, there exists a need for the specific detection of small RNA molecules, such as miRNA molecules, in complex mixtures or multiplex reactions.

SUMMARY OF THE INVENTION

[08] Compositions and methods of the invention provide a solution for the long felt need for the specific detection of small RNA molecules, such as miRNA molecules, in complex mixtures or multiplex reactions. This can be achieved by ligation of any small RNA molecule to a unique, sequence-specific tag molecule at a single temperature by normalizing the melting temperature of hybridization between the target small RNA and a bridge molecule which directs the specific attachment of a unique tag. The ligated tags subsequently normalize the melting temperatures of the mixed population of tagged small RNAs, allowing them to be subjected to a multiplexed hybridization assay in which it is possible to distinguish small RNAs in a sequence-specific manner at the same temperature.

[09] Specifically, the invention provides a composition including: (a) a tag, wherein the tag includes a first DNA sequence and a reporter attachment region to which are attached one or more reporter molecules that produce a signal and wherein the first DNA sequence includes an alien DNA sequence; and (b) a bridge, wherein the bridge includes a second DNA sequence that is complementary to a RNA molecule and a third DNA sequence that is complementary to the first DNA sequence of the tag.

[10] In one aspect of this composition the second DNA sequence is complementary to a RNA molecule or a portion thereof. A portion of the RNA molecule is less than the entire length of the RNA molecule. In certain aspects the portion is 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 99%, or any percentage point in between of the entire length of the RNA molecule. Moreover, the portion is either continuous or discontinuous.

[11] In another aspect of the composition, the first DNA sequence of the tag includes alien

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1287

DNA, wherein the alien DNA does not cross-hybridize with any other DNA or RNA molecule in the composition other than the third DNA sequence of the bridge. Alternatively, or in addition, the alien DNA cross-hybridizes with less than 15% of DNA and RNA molecules in the composition other than the third DNA sequence of the bridge. Specifically, the alien DNA sequence cross-hybridizes with no greater than 15%, 10%, 5%, 2%, 1% or any percentage point in between, of DNA and RNA molecules in the composition other than the third DNA sequence of the bridge. In certain aspects of the composition, the tag contains a first DNA sequence composed of alien DNA, which has no greater than 85% identity across 35 bases with a maximum continuous homology of 15 bases or less to any known DNA or RNA sequence in the composition. Alternatively, the alien DNA has no greater than 85% identity across 35 bases with a maximum continuous homology of 15 bases or less to any known genomic DNA or RNA sequence. Moreover, the alien DNA has no greater than 85% identity across 35 bases with a maximum continuous homology of 15 bases or less to any known DNA or RNA sequence from or transcribed by a genome from which the target RNA molecule is transcribed. Target RNA molecules are from or transcribed by the genomes of any species, including, but not limited to, animals, plants, bacteria, fungi and viruses. Preferred species are mammals, including primates, preferably humans and rodents including rabbit, rat and mice.

[12] In certain aspects of the tag, the first DNA sequence and the reporter attachment region are tangential or overlapping. The first DNA sequence and the reporter attachment region may overlap by a number of DNA nucleotides ranging from 1-5, 5-10, 10-15, 15-20, 20-25, 25-30, 30-35, 35-40, 40-45, 45-50, 50-55, 55-60, 60-65, 65-70, 75-80, 80-85, 85-90, 90-95, 95-100, or any number of nucleotides in between. Moreover, the first DNA sequence and the reporter attachment region may overlap by a percentage of the total number of nucleotides present in the first DNA sequence ranging from 1-5%, 5-10%, 10-15%, 15-20%, 20-25%, 25-30%, 30-35%, 35-40%, 40-45%, 45-50%, 50-55%, 55-60%, 60-65%, 65-70%, 75-80%, 80-85%, 85-90%, 90-95%, 95-100%, or any percentage of nucleotides of the first DNA sequence in between. In a particular aspect, the reporter attachment region may be entirely contained within the first DNA sequence.

[13] In one embodiment of the composition, the second and third DNA sequences of the bridge are contiguous. Alternatively, the second and third DNA sequences are connected by a linker molecule containing at least two deoxyribonucleotides. In certain aspects of the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1288

composition, the linker molecule contains between 2-100 deoxyribonucleotides, and specifically, between 2-10 deoxyribonucleotides, 10-20 deoxyribonucleotides, 20-30 deoxyribonucleotides, 30-40 deoxyribonucleotides, 40-50 deoxyribonucleotides, 50-60 deoxyribonucleotides, 60-70 deoxyribonucleotides, 70-80 deoxyribonucleotides, 80-90 deoxyribonucleotides, 90-100 deoxyribonucleotides, or any number of deoxyribonucleotides in between. The linker molecule may contain deoxyribonucleotide (DNA) sequence obtained or derived from any species, including alien or synthetic sequences.

[14] In certain aspects of the composition, the second DNA sequence of the bridge and the RNA molecule form a DNA/RNA heteroduplex having a melting temperature of between 37-95°C. Alternatively, the second DNA sequence of the bridge and the RNA molecule form a DNA/RNA heteroduplex having a melting temperature of between 44-53°C. In other aspects, the third DNA sequence of the bridge and the first sequence of the tag form a DNA/DNA duplex having a melting temperature of between 37-95°C. Alternatively, the third DNA sequence of the bridge and the first sequence of the tag form a DNA/DNA duplex having a melting temperature of between 44-53°C.

[15] In a preferred embodiment of the composition, the second and third DNA sequences of the bridge form nucleic acid duplexes having substantially the same melting temperature. In certain aspects, the second and third DNA sequences of the bridge form nucleic acid duplexes having melting temperatures within 1°C, 2°C, 5°C, 10°C, 15°C, 20°C, or any degree in between of one another. Preferably, the second and third DNA sequences of the bridge form nucleic acid duplexes having melting temperatures within 1°C, 2°C, or 5°C of one another.

[16] In another embodiment of the composition, the third DNA sequence includes an alien DNA sequence. In one aspect of this embodiment, the third DNA sequence includes a portion of an alien DNA sequence.

[17] In another embodiment of the composition, the RNA comprises a non-coding RNA. Exemplary non-coding RNA include, but are not limited to, a transfer RNA (tRNA), a small nucleolar RNA (snoRNA), a microRNA (miRNA), a short interfering RNA (siRNA), a short hairpin RNA (shRNA), a repeat associated small interfering RNA (rasiRNA), and a piwi-interacting RNA (piRNA). In a preferred aspect, the non-coding RNA is a miRNA. Alternatively or in addition, the RNA molecule includes between 10-300 ribonucleotides.

[18] In certain embodiments of the composition, the RNA molecule and/or the tag contain

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1289

sequences that specifically hybridize to the bridge with complete complementarity. Alternatively, or in addition, the RNA molecule and/or the tag contain sequences that specifically hybridize to the bridge with partial complementarity. In certain aspects, partial complementarity is defined as less than complete or less than 100% complementarity. Alternatively, partial complementarity is defined as binding with less than 100%, 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20%, 10%, or any percentage value in between.

[19] The invention further provides a kit containing: (a) a composition including: (1) a tag, wherein the tag includes a first DNA sequence and a reporter attachment region to which are attached one or more reporter molecules that produce a signal; and (2) a bridge, wherein the bridge includes a second DNA sequence that is complementary to a RNA molecule and a third DNA sequence that is complementary to the first DNA sequence of the tag; and (b) a substance selected from the group consisting of a volume excluder and a nuclease. Optionally, the kit further includes a ligase.

[20] In one embodiment of the kit, the selected substance is a volume excluder. A nonlimiting example of a preferred volume excluder is polyethylene glycol (PEG).

[21] In another embodiment of the kit, the selected substance is a nuclease. The nuclease is preferably a DNA-specific exonuclease. A nonlimiting example of a DNA-specific exonuclease is a lambda exonuclease.

[22] In certain embodiments of the kit, the second and third DNA sequences of the bridge are contiguous. Alternatively, the second and third DNA sequences are connected by a linker molecule. In certain aspects of the kit, the linker molecule contains between 2-100 deoxyribonucleotides, and specifically, between 2-10 deoxyribonucleotides, 10-20 deoxyribonucleotides, 20-30 deoxyribonucleotides, 30-40 deoxyribonucleotides, 40-50 deoxyribonucleotides, 50-60 deoxyribonucleotides, 60-70 deoxyribonucleotides, 70-80 deoxyribonucleotides, 80-90 deoxyribonucleotides, 90-100 deoxyribonucleotides, or any number of deoxyribonucleotides in between. The linker molecule may contain deoxyribonucleotide (DNA) sequence obtained or derived from any species, including alien sequences.

[23] In other embodiments of the kit, the second DNA sequence of the bridge and the RNA molecule form a DNA/RNA heteroduplex having a melting temperature of between 37-95°C. Alternatively, the second DNA sequence of the bridge and the RNA molecule form a DNA/RNA heteroduplex having a melting temperature of between 44-53°C. In a further embodiment, the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1290

third DNA sequence of the bridge and the first sequence of the tag form a DNA/DNA duplex having a melting temperature of between 37-95°C. Alternatively, the third DNA sequence of the bridge and the first sequence of the tag form a DNA/DNA duplex having a melting temperature of between 44-53°C.

[24] In a preferred embodiment, the second and third DNA sequences of the bridge form nucleic acid duplexes having substantially the same melting temperature. In certain aspects, the second and third DNA sequences of the bridge form nucleic acid duplexes having melting temperatures within 1°C, 2°C, 5°C, 10°C, 15°C, 20°C, or any degree in between of one another. Preferably, the second and third DNA sequences of the bridge form nucleic acid duplexes having melting temperatures within 1°C, 2°C, or 5°C of one another.

[25] In another preferred embodiment, the third DNA sequence includes an alien DNA sequence.

[26] In certain embodiments of the kit, the RNA molecule and/or the tag contain sequences that specifically hybridize to the bridge with complete complementarity. Alternatively, or in addition, the RNA molecule and/or the tag contain sequences that specifically hybridize to the bridge with partial complementarity. In certain aspects, partial complementarity is defined as less than complete or less than 100% complementarity. Alternatively, partial complementarity is defined as binding with less than 100%, 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20%, or 10% complementarity or any percentage value in between.

[27] The invention further provides a method of detecting a RNA molecule including: (a) providing a sample containing at least one RNA molecule; (b) providing a tag, wherein the tag includes a first DNA sequence and a reporter attachment region to which are attached one or more reporter molecules that produce a signal; (c) providing a bridge, wherein the bridge includes a second DNA sequence that is complementary to the RNA molecule and a third DNA sequence that is complementary to the first DNA sequence of the tag; and (d) providing a buffer; (e) specifically annealing the RNA molecule, bridge, and tag at between 37-95°C; (f) holding the annealed reaction at between 37-95°C; (g) providing a ligase buffer; (h) providing a ligase directly to annealed reaction at between 37-95°C; (i) ligating the RNA molecule to the tag at one or more temperatures between 37-95°C; and (j) detecting the signal.

[28] In certain embodiments, this method of detecting a RNA molecule further includes the step of denaturing the RNA molecule, bridge, and tag prior to specifically annealing the RNA

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1291

molecule, bridge, and tag at between 37-95°C.

[29] In another embodiment of this method, steps (e), (f), (h), and (i) are performed at between 43-52°C.

[30] In other embodiments, this method of detecting a RNA molecule further includes providing a volume excluder after holding the annealed reaction at between 37-95°C. A nonlimiting example of a volume excluder is polyethylene glycol (PEG).

[31] In certain embodiments, the method further includes providing a nuclease to the ligation reaction after ligating the RNA molecule to the tag at between 37-95°C. In one aspect, the nuclease is a DNA-specific exonuclease. A nonlimiting example of the DNA-specific exonuclease is a lambda exonuclease.

[32] In a preferred embodiment of this method of detecting a RNA molecule, the second and third DNA sequences of the bridge are contiguous. Alternatively, the second and third DNA sequences are connected by a linker molecule. In certain aspects of this method, the linker molecule contains between 2-100 deoxyribonucleotides, and specifically, between 2-10 deoxyribonucleotides, 10-20 deoxyribonucleotides, 20-30 deoxyribonucleotides, 30-40 deoxyribonucleotides, 40-50 deoxyribonucleotides, 50-60 deoxyribonucleotides, 60-70 deoxyribonucleotides, 70-80 deoxyribonucleotides, 80-90 deoxyribonucleotides, 90-100 deoxyribonucleotides, or any number of deoxyribonucleotides in between. The linker molecule may contain deoxyribonucleotide (DNA) sequence obtained or derived from any species, including alien sequences.

[33] In another embodiment of this method of detecting a RNA molecule, the RNA comprises a non-coding RNA. Exemplary non-coding RNA include, but are not limited to, a transfer RNA (tRNA), a small nucleolar RNA (snoRNA), a microRNA (miRNA), a short interfering RNA (siRNA), a short hairpin RNA (shRNA), a repeat associated small interfering RNA (rasiRNA), and a piwi-interacting RNA (piRNA). In a preferred aspect of this method, the non-coding RNA is a miRNA.

[34] In certain embodiments of this method of detecting a RNA molecule, the second DNA sequence of the bridge and the RNA molecule form a DNA/RNA heteroduplex having a melting temperature of between 37-95°C. Alternatively, the second DNA sequence of the bridge and the RNA molecule form a DNA/RNA heteroduplex having a melting temperature of between 44-53°C. In other embodiments, the third DNA sequence of the bridge and the first sequence of the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1292

tag form a DNA/DNA duplex having a melting temperature of between 37-95°C. Alternatively, the third DNA sequence of the bridge and the first sequence of the tag form a DNA/DNA duplex having a melting temperature of between 44-53°C.

[35] In a preferred embodiment of this method of detecting a RNA molecule, the second and third DNA sequences of the bridge form nucleic acid duplexes having substantially the same melting temperature. In certain aspects, the second and third DNA sequences of the bridge form nucleic acid duplexes having melting temperatures within 1°C, 2°C, 5°C, 10°C, 15°C, 20°C, or any degree in between of one another. Preferably, the second and third DNA sequences of the bridge form nucleic acid duplexes having melting temperatures within 1°C, 2°C, or 5°C of one another.

[36] In one embodiment of this method of detecting a RNA molecule, the third DNA sequence comprises an alien DNA sequence.

[37] In certain aspects of this method of detecting a RNA molecule, the RNA molecule and/or the tag contain sequences that specifically hybridize to the bridge with complete complementarity. Alternatively, or in addition, the RNA molecule and/or the tag contain sequences that specifically hybridize to the bridge with partial complementarity. In certain aspects, partial complementarity is defined as less than complete or less than 100% complementarity. Alternatively, partial complementarity is defined as binding with less than 100%, 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20%, or 10% complementarity, or any percentage value in between.

[38] The invention further provides a method of multiplex detection of a plurality of RNA molecules including: (a) providing a sample containing a plurality of RNA molecules; (b) providing a plurality of tags, wherein each tag includes a first DNA sequence and at least one reporter attachment region to which are attached one or more reporter molecules that produce a signal; (c) providing a plurality of bridges, wherein each bridge includes a second DNA sequence that is complementary to one RNA molecule and a third DNA sequence that is complementary to the first DNA sequence of one tag, wherein each bridge specifically anneals to one species of RNA molecule and one species of tag, wherein the one species of tag produces a signal that differentially labels one species of RNA molecule compared to other species of RNA molecules when it is joined to the tag; (d) providing a buffer; (e) specifically annealing the RNA molecules, bridges, and tags at between 37-95°C; (f) holding the annealed reaction at between

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1293

37-95°C; (g) providing a ligase buffer; (h) providing a ligase directly to annealed reaction at between 37-95°C; (i) ligating the RNA molecules, bridges, and tags at one or more temperatures between 37-95°C; and (j) detecting one or more signals.

[39] In one embodiment, this method of multiplex detection of a plurality of RNA molecules further includes the step of denaturing the RNA molecules, bridges, and tags prior to specifically annealing the RNA molecules, bridges, and tags at between 37-95°C.

[40] In another embodiment of this method of multiplex detection of a plurality of RNA molecules, steps (e), (f), (h), and (i) are performed at between 43-52°C.

[41] In certain embodiments, this method of multiplex detection of a plurality of RNA molecules further includes providing a volume excluder after holding the annealed reaction at between 37-95°C. A nonlimiting example of the volume excluder is polyethylene glycol (PEG). Alternatively, or in addition, the method further includes providing a nuclease to the ligation reaction after ligating the RNA molecule to the tag at between 37-95°C. In one aspect, the nuclease is a DNA-specific exonuclease. A nonlimiting example of the DNA-specific exonuclease is a lambda exonuclease.

[42] In a preferred embodiment of this method of multiplex detection of a plurality of RNA molecules, the second and third DNA sequences of the bridges are contiguous. Alternatively, the second and third DNA sequences are connected by a linker molecule. In certain aspects of this method, the linker molecule contains between 2-100 deoxyribonucleotides, and specifically, between 2-10 deoxyribonucleotides, 10-20 deoxyribonucleotides, 20-30 deoxyribonucleotides, 30-40 deoxyribonucleotides, 40-50 deoxyribonucleotides, 50-60 deoxyribonucleotides, 60-70 deoxyribonucleotides, 70-80 deoxyribonucleotides, 80-90 deoxyribonucleotides, 90-100 deoxyribonucleotides, or any number of deoxyribonucleotides in between. The linker molecule may contain deoxyribonucleotide (DNA) sequence obtained or derived from any species, including alien sequences.

[43] In another preferred embodiment, the third DNA sequence includes an alien DNA sequence.

[44] In another embodiment of this method of multiplex detection of a plurality of RNA molecules, the RNA molecules comprise a non-coding RNA. Nonlimiting examples of the non-coding RNA include, but are not limited to, a transfer RNA (tRNA), a small nucleolar RNA (snoRNA), a microRNA (miRNA), a short interfering RNA (siRNA), a short hairpin RNA

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1294

(shRNA), a repeat associated small interfering RNA (rasiRNA), or a piwi-interacting RNA (piRNA). Preferably, the non-coding RNA is a miRNA.

[45] In certain embodiments of this method of multiplex detection of a plurality of RNA molecules, the second DNA sequence of the bridges and the RNA molecules form a DNA/RNA heteroduplex having a melting temperature of between 37-95°C. Alternatively, the second DNA sequence of the bridges and the RNA molecules form a DNA/RNA heteroduplex having a melting temperature of between 44-53°C. In other embodiments, the third DNA sequence of the bridges and the first sequence of the tags form a DNA/DNA duplex having a melting temperature of between 37-95°C. Alternatively, the third DNA sequence of the bridges and the first sequence of the tags form a DNA/DNA duplex having a melting temperature of between 44-53°C.

[46] In a preferred embodiment of this method of multiplex detection of a plurality of RNA molecules, the second and third DNA sequences of the bridge form nucleic acid duplexes having substantially the same melting temperature. In certain aspects, the second and third DNA sequences of the bridge form nucleic acid duplexes having melting temperatures within 1°C, 2°C, 5°C, 10°C, 15°C, 20°C, or any degree in between of one another. Preferably, the second and third DNA sequences of the bridge form nucleic acid duplexes having melting temperatures within 1°C, 2°C, or 5°C of one another.

[47] In one embodiment of this method of multiplex detection of a plurality of RNA molecules, the RNA molecules and/or the tags contain sequences that specifically hybridize to the bridge with complete complementarity. Alternatively, or in addition, the RNA molecules and/or the tags contain sequences that specifically hybridize to the bridge with partial complementarity. In certain aspects, partial complementarity is defined as less than complete or less than 100% complementarity. Alternatively, partial complementarity is defined as binding with less than 100%, 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20%, 10%, or any percentage value in between.

[48] The invention also provides a method of making a nucleic acid bridge molecule, including: (a) selecting a RNA molecule; (b) selecting a segment of the RNA molecule that forms a DNA/RNA heteroduplex with a DNA molecule that specifically hybridizes to the segment of the RNA molecule having a melting temperature of between 37-95°C; (c) generating a first DNA bridge molecule that specifically hybridizes to the segment of the RNA molecule

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1295

having a melting temperature of between 37-95°C; (d) selecting a tag, wherein the tag includes a DNA sequence that is an alien sequence; (e) selecting a segment of the tag that forms a DNA/ DNA duplex with a DNA molecule that specifically hybridizes to the segment of the tag having a melting temperature of between 37-95°C; (f) generating a second DNA bridge molecule that specifically hybridizes to the segment of the tag having a melting temperature of between 37-95°C; and (g) uniting the first DNA bridge molecule with the second DNA bridge molecule, thereby forming the nucleic acid DNA bridge molecule that specifically hybridizes to the segment of the target RNA molecule having a melting temperature of between 37-95°C and the segment of the tag having a melting temperature of between 37-95°C.

[49] In certain embodiments of this method of making a nucleic acid bridge molecule, the uniting step includes ligating the first DNA bridge molecule to the second DNA bridge molecule. Alternatively, the uniting step includes synthesizing a nucleic acid DNA bridge molecule containing the sequence of the first DNA bridge molecule and the sequence of the second DNA bridge molecule.

[50] In a preferred embodiment of this method of making a nucleic acid bridge molecule, the first DNA bridge molecule and the second DNA bridge molecule have substantially the same melting temperature. In certain aspects, the second and third DNA sequences of the bridge form nucleic acid duplexes having melting temperatures within 1°C, 2°C, 5°C, 10°C, 15°C, 20°C, or any degree in between of one another. Preferably, the second and third DNA sequences of the bridge form nucleic acid duplexes having melting temperatures within 1°C, 2°C, or 5°C of one another.

[51] In another preferred embodiment of this method of making a nucleic acid bridge molecule, the second DNA bridge molecule is selected from the 5' end of the first DNA sequence of the tag. Specifically, the second DNA molecule is selected from a portion of the first DNA sequence of the tag that includes between 2-5 nucleotides, 2-10 nucleotides, 2-20 nucleotides, 2-30 nucleotides, 2-40 nucleotides, 2-50 nucleotides, 2-60 nucleotides, 2-70 nucleotides, 2-80 nucleotides, 2-90 nucleotides, 2-100 nucleotides or any range in between of nucleotides from the 5' end of the first DNA sequence of the tag. Preferably, the second DNA molecule is selected from a portion of the first DNA sequence of the tag that includes between 2-10 nucleotides or between 2-20 nucleotides from the 5' end of the first DNA sequence of the tag.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1296

[52] In particular embodiments of this method of making a nucleic acid bridge molecule, the DNA/RNA heteroduplex has a melting temperature of between 43-52°C. Alternatively, or in addition, the DNA/ DNA duplex has a melting temperature of between 43-52°C.

[53] In a preferred embodiment of this method of making a nucleic acid bridge molecule, the first DNA bridge molecule and the second DNA bridge molecule are contiguous. Alternatively, the second and third DNA sequences are connected by a linker molecule. In certain aspects of this method, the linker molecule contains between 2-100 deoxyribonucleotides, and specifically, between 2-10 deoxyribonucleotides, 10-20 deoxyribonucleotides, 20-30 deoxyribonucleotides, 30-40 deoxyribonucleotides, 40-50 deoxyribonucleotides, 50-60 deoxyribonucleotides, 60-70 deoxyribonucleotides, 70-80 deoxyribonucleotides, 80-90 deoxyribonucleotides, 90-100 deoxyribonucleotides, or any number of deoxyribonucleotides in between. The linker molecule may contain deoxyribonucleotide (DNA) sequence obtained or derived from any species, including alien sequences.

[54] In other embodiments of this method of making a nucleic acid bridge molecule, the RNA molecule is a miRNA. Alternatively, the RNA molecule includes a non-coding RNA. Preferably, the non-coding RNA is a miRNA.

[55] Particularly in those embodiments of this method of making a nucleic acid bridge molecule wherein the RNA molecule is an miRNA, but with respect to all embodiments presented herein, the first DNA bridge molecule hybridizes to the RNA molecule with complete complementarity. Alternatively, the first DNA bridge molecule hybridizes to the RNA molecule with partial complementarity. With respect to all embodiments presented herein, the second DNA bridge molecule hybridizes to the tag with complete complementarity. Alternatively, the second DNA bridge molecule hybridizes to the tag with partial complementarity. In certain aspects, partial complementarity is defined as less than complete or less than 100% complementarity. Alternatively, partial complementarity is defined as binding with less than 100%, 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20%, 10%, or any percentage value in between.

BRIEF DESCRIPTION OF THE DRAWINGS

[56] Figure 1 is a graph depicting human miRNA melting temperature (Tm) distribution as the percentage (%) of human miRNAs versus Tm in degrees Celsius (°C).

[57] Figure 2 is a schematic representation of the chimeric RNA:DNA heteroduplex, where

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1297

the reporter attachment region (A) comprises the DNA tag and the DNA sequences (C) and (D) comprise the bridge.

[58] Figure 3 is a schematic representation of the chimeric RNA:DNA heteroduplex, where the reporter attachment region (A) and an additional sequence (B) comprise the DNA tag and the DNA sequences (C) and (D) comprise the bridge.

[59] Figure 4 is a photograph of a gel electrophoresis assay. The results of successful ligation reactions completed at temperatures ranging from 45-52 °C and containing miRNAs, tags and the corresponding bridges are shown.

[60] Figure 5A is a graph depicting miRNA expression levels in total RNA isolated from skeletal muscle versus total RNA isolated from lung tissue. Samples were measured using the nCounter® miRNA Expression Assay Kit.

[61] Figure 5B is a graph depicting miRNA expression levels in total RNA isolated from heart tissue versus total RNA isolated from colon tissue. Samples were measured using the nCounter® miRNA Expression Assay Kit.

[62] Figure 5C is a graph depicting miRNA expression levels in total RNA isolated from heart tissue versus total RNA isolated from lung tissue. Samples were measured using the nCounter® miRNA Expression Assay Kit.

[63] Figure 5D is a graph depicting miRNA expression levels in total RNA isolated from colon tissue versus total RNA isolated from skeletal muscle. Samples were measured using the nCounter® miRNA Expression Assay Kit.

[64] Figure 6A is a graph depicting the expression levels of human miRNAs miR-133a, miR-143, miR-16, miR-21, miR-29a, and miR-30 in total RNA isolated from skeletal muscle, lung, colon, and heart tissue. Samples were measured using the nCounter® miRNA Expression Assay Kit.

[65] Figure 6B is a graph depicting the expression levels of human miRNA miR-1in total RNA isolated from skeletal muscle, lung, colon, and heart tissue. Samples were measured using the nCounter® miRNA Expression Assay Kit.

[66] Figure 7 is a graph depicting the results of multiplexed nCounter® miRNA Expression Assays for 55 human miRNAs run on 100 ng of total RNA purified from either formalin-fixed, paraffin-embedded (FFPE) or frozen human liver samples collected from the same tissue source. The results are displayed as counts in matched FFPE Liver RNA versus counts in matched

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1298

frozen liver control.

[67] Figure 8 is a graph depicting the correlation of two technical replicates in which multiplexed nCounter® miRNA Expression Assays for 676 human miRNAs were run on 100 ng of human brain total RNA. The data are displayed as nCounter® counts (Technical Replicate #2) versus nCounter® counts (Technical Replicate #1).

[68] Figure 9A is a series of graphs depicting mixtures of five canonical (left), variant 1 (center), or variant 2 (right) synthetic miRNAs assayed with the canonical and variant bridge pools in multiplexed nCounter® miRNA Expression Assays. The graphs display the counts resulting when each of the three mixtures was assayed the canonical (black bars), variant 1 (white bars), and variant 2 (gray bars) bridge pools.

[69] Figure 9B is a graph depicting results from nCounter® miRNA Expression Assays used to measure a mixture containing 60% canonical (black), 30% variant 1 (white), and 10% variant 2 (gray) for each of 5 miRNAs. The relative abundance of each species measured in the assay is indicated as a percentage of total signal for each miRNA.

DETAILED DESCRIPTION

[70] The invention provides a sensitive, hybridization-based technology for the multiplexed analysis of mRNA gene expression. In certain embodiments of the invention, the nCounter® Analysis System is used. In mRNA detection experiments using the nCounter® Analysis System, high hybridization reaction temperatures (typically 65°C) promote the specific target/reporter interaction of a 70-100 base target region. Such assays do not allow for the detection of small nucleic acid species, *e.g.* miRNAs, due to the low melting temperature (T_m) of such short sequences. Thus, the invention solves this problem by providing a novel assay, that may be used in the nCounter® miRNA Expression Assay Kit, which is compatible with the nCounter® Analysis System described herein and allows for specific and sensitive detection of short RNAs, using microRNAs (miRNAs) as a model system. The assay involves attaching a specific DNA tag to each discrete species of small RNA, thereby creating a chimeric RNA-DNA target with a T_m that is high enough to be compatible with the standard hybridization conditions of the nCounter® Analysis System. The addition of the tag can be done in a multiplexed fashion, such that a single sample can be assayed for hundreds of species of small RNAs simultaneously. A key feature of this invention is the highly specific attachment of each tag

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1299

under narrowly defined reaction conditions, such that the specificity and efficiency of the tag attachment reaction is retained in a multiplex format where all of the separate interactions are taking place simultaneously and under the same reaction conditions.

[71] MicroRNAs (miRNAs) are small, endogenous, non-coding, regulatory, single-stranded RNA molecules of about 21-23 nucleotides in length. While miRNAs in general exhibit a great deal of sequence diversity, small groups of miRNAs (miRNA families) exist with >90% sequence identity. Hybridization-based assays for specific detection of nearly identical species are most sensitive when hybridization occurs near the T_m of the target species. Such conditions greatly reduce cross-hybridization between nearly identical sequences by maximizing thermodynamic destabilization between target and mismatched species. The overall sequence diversity of miRNAs and other small RNAs complicates specific multiplexed detection of their expression by hybridization because it means that the optimal hybridization temperature (*i.e.* target T_m) varies from target to target (Figure 1). A second key feature of the nCounter® miRNA Expression Assay Kit is the target-specific design of each synthetic DNA tag which allows for the normalization of the diverse T_m's of naturally-occurring small RNA sequences; in other words, the individual tags are designed to make the T_m's of the final RNA-DNA chimeras equivalent, regardless of the T_m of the underlying biological RNA sequence. This allows the subsequent hybridization detection assay to be done at a single temperature for all the diverse species, and thus allows for multiplexing of the detection reaction.

MicroRNAs

[72] MicroRNAs (miRNAs) are small, non-coding regulatory RNAs that are hypothesized to act by inhibiting translation of messenger RNA (mRNA) into protein by binding to the 3' untranslated region (UTR) of their target mRNAs. MiRNAs inhibit mRNA translation by either causing mRNA degradation or inhibiting translation itself.

[73] MiRNAs are single-stranded RNA molecules of about 21-23 nucleotides in length. MiRNAs are encoded by endogenous and exogenous genes that are transcribed from genomic DNA; however, miRNAs are typically not translated into polypeptide sequences. As such, miRNAs are considered in the art as "non-coding RNAs." The term "endogenous" gene as used herein is meant to encompass all genes that naturally occur within the genome of an individual. The term "exogenous" gene as used herein is meant to encompass all genes that do not naturally occur within the genome of an individual. For example, a miRNA could

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1300

be introduced exogenously by a virus.

[74] MiRNA molecules of the invention include pri-miRNA, pre-miRNA, and mature miRNAs. Mature miRNAs are initially transcribed from genomic DNA into pri-miRNA. Pri-miRNA contains at least 1 (up to 6 when transcribed from polycistronic units) hairpin loop structures that are each approximately 70 nucleotides in length. There is a potential for a single pri-miRNA to contain multiple miRNAs. Pri-miRNA can also be subject to RNA editing wherein the miRNA processing or specificity is altered. Pri-miRNA is cleaved in the cell nucleus about 11 nucleotides from the base of each hairpin structure (2 helical RNA turns into the stem) to form pre-miRNA. Staggered cuts are introduced into the ends of the hairpin loop arms resulting in a 2 nucleotide overhang on the 3' end and a phosphate on the 5' end to produce a pre-miRNA of approximately 70 nucleotides in length. In the cytoplasm the pre-miRNA is cleaved and results in an imperfect miRNA:miRNA* duplex of approximately 20-25 nucleotide pairs in length that contains the mature miRNA strand and its opposite complementary miRNA* strand. The passenger strand, or complementary miRNA strand*, is normally degraded and present in lower levels in cells in the steady state, although there are instances where both strands of the duplex become functional miRNAs that target different mRNA populations.

[75] The term "miRNA" is equivalent to other terms used to describe nucleic acid molecules that are capable of mediating sequence-specific gene silencing or interference, *e.g.*, short interfering RNA (siRNA), double-stranded RNA (dsRNA), interfering RNA (RNAi), short hairpin RNA (shRNA), short interfering oligonucleotides, short interfering nucleic acid, short interfering chemically-modified siRNA, post-transcriptional gene silencing RNA (ptgsRNA), and other art-recognized equivalents. As used herein, the term "gene silencing" is meant to describe the downregulation, knock-down, degradation, inhibition, suppression, repression, prevention, or decreased expression of a gene, transcript and/or polypeptide product. Gene silencing and interference also describe the prevention of translation of mRNA transcripts into a polypeptide. Translation is prevented, inhibited, or decreased by degrading mRNA transcripts or blocking mRNA translation.

[76] The bridges of the invention bind specifically to their target miRNAs based on either partial or complete sequence complementarity. Exemplary miRNAs are provided in Table 1, below.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1301

[77] Table 1: Exemplary Human miRNAs

SEQ ID NO:	Sequence	miRBase™ ID
1	UGAGGUAGUAGGUUGUAUAGUU	hsa-let-7a
2	UGAGGUAGUAGGUUGUGUGGUU	hsa-let-7b
3	UGAGGUAGUAGGUUGUAUGGUU	hsa-let-7c
4	AGAGGUAGUAGGUUGCAUAGU	hsa-let-7d
5	UGAGGUAGGAGGUUGUAUAGU	hsa-let-7e
6	UGAGGUAGUAGAUUGUAUAGUU	hsa-let-7f
7	UGAGGUAGUAGUUUGUACAGU	hsa-let-7g
8	UGAGGUAGUAGUUUGUGCUGU	hsa-let-7i
9	UGGAAUGUAAAGAAGUAUGUA	hsa-miR-1
10	AACCCGUAGAUCCGAACUUGUG	hsa-miR-100

All miRNA sequences are publicly available at (www.mirbase.org).

nCounter® Analysis System Protocol overview and Data format

[78] The basis of the nCounter® Analysis system is the unique code assigned to each gene to be assayed (International Patent Application No. PCT/US2008/059959 and Geiss *et al.* Nature Biotechnology. 2008. 26(3): 317-325; the contents of which are each incorporated herein by reference in their entireties). The code is composed of an ordered series of colored fluorescent spots which create a unique barcode for each target to be assayed. A pair of probes is designed for each target, a biotinylated capture probe and a reporter probe carrying the fluorescent barcode.

[79] Specific reporter and capture probes are synthesized for each target. Briefly, sequence-specific DNA oligonucleotide probes are attached to code-specific reporter molecules. Capture probes are made by ligating a second sequence-specific DNA oligonucleotide for each target to a universal oligonucleotide containing biotin. Reporter and capture probes are all pooled into a single hybridization mixture, the "probe library".

[80] The expression levels of all targets are measured in a single multiplexed hybridization reaction. The sample is combined with the probe library, and hybridization occurs in solution. After hybridization, the tripartite hybridized complexes are purified in a two-step procedure using magnetic beads linked to oligonucleotides complementary to universal sequences present on the capture and reporter probes. This dual purification process allows the hybridization reaction to be driven to completion with a large excess of target-specific probes, as they are ultimately removed, and, thus, do not interfere with binding and imaging of the sample. All post

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1302

hybridization steps are handled robotically on a custom liquid-handling robot (Prep Station, NanoString Technologies).

[81] Purified reactions are deposited by the Prep Station into individual flow cells of a sample cartridge, bound to a streptavidin-coated surface via the capture probe, electrophoresed to elongate the reporter probes, and immobilized. After processing, the sample cartridge is transferred to a fully automated imaging and data collection device (Digital Analyzer, NanoString Technlogies). The expression level of a target is measured by imaging each sample and counting the number of times the code for that target is detected. For each sample, typically 600 fields-of-view (FOV) are imaged (1376 X 1024 pixels) representing approximately 10 mm^2 of the binding surface. Typical imaging density is 100-1200 counted reporters per field of view depending on the degree of multiplexing, the amount of RNA, and overall target expression levels. Data is output in simple spreadsheet format listing the number of counts per target, per sample.

nCounter® miRNA Expression Assay Sample Preparation Protocol

[82] In the nCounter® miRNA Expression Assay, the miRNA samples must be tagged before they can be introduced into the probe hybridization reaction of the nCounter® Analysis System described above. Unless otherwise noted, the data in the examples provided were generated using an embodiment of the nCounter® miRNA Expression Assay Kit Sample Preparation Protocol described below, followed by the standard nCounter® Assay Protocol described briefly above and in Geiss *et al.*

[83] Annealing buffer (200mM Tris, 750 mM KCl) was mixed with a tag pool. The tag pool includes 60 nM each of a set of the synthetic DNA oligonucleotides (Integrated DNA Technologies, Coralville, IA) similar to those described in Table 2. The exact composition of this pool depends on the specific miRNAs being measured. An equivalent amount of a bridge pool is added to the mixture. The bridge pool consists of 55 nM each of a set of synthetic DNA oligonucleotides (Integrated DNA Technologies) similar to those described in Table 3. The exact composition of this pool depended on the specific miRNAs being measured, and matched the profile of the tag pool. 100 ng of total RNA or a defined amount of synthetic RNA target (Integrated DNA Technologies) is added and DEPC-treated water (Ambion, Austin, TX) was added to normalize the volume of each reaction. Samples were denatured and annealed in a thermocycler (BioRad, Hercules, CA) °C. An equal-parts mixture of 50% PEG-4000

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1303

(CalBioChem/EMD Biosciences, San Diego, CA) and T4 DNA Ligase buffer (Enzymatics, Beverly, MA) was prepared for addition. Samples were briefly removed from the thermocycler, PEG/ligase buffer stock solution was added to each tube, and samples were mixed and spun down. Tubes were returned to the thermocycler and T4 DNA Ligase (Enzymatics) was added directly to the samples without removing them from the thermocycler. All thermocycler reactions were performed at temperatures in the range of 45-50 °C (Figure 4).

[84] Lambda Exonuclease (Enzymatics) was added to each sample at room temperature. Samples were incubated at 37°C for 2 hours, heat-inactivated at 70°C and chilled. Sample is diluted 10-fold with DEPC-treated water and used as the substrate in a nCounter® hybridization reaction using a Codeset for miRNA detection (NanoString Technologies, Seattle, WA), following the manufacturer's instructions.

[85] Loading of ligase should take approximately 6 minutes per 24 samples. Direct addition of ligase to the reaction tubes on the heat block at 48°C is a critical step. Failure to do so will result in decreased assay performance, especially in terms of assay specificity.

[86] The denaturing, annealing, ligation, and purification reactions described herein may be performed within a broader range of temperatures that span, for example, from 37-95°C. Specifically, the denaturing, annealing, ligation, and purification reactions described herein are performed with one or more of the following ranges including, but limited to, 37-45°C, 44-53°C, 45-52 °C, 45-50°C, 50-55°C, 55-60°C, 65-70°C, 70-75°C, 75-80°C, 85-90°C, and 90-95°C. Alternatively, the denaturing, annealing, ligation, and purification reactions described herein are performed with one or more of the following ranges including, but limited to, 37-45 °C, 37-50 °C, 37-55 °C, 37-60 °C, 37-65 °C, 37-70 °C, 37-75 °C, 37-80 °C, 37-85 °C, 37-90 °C, and 37-95 °C.

[87] nCounter® In the examples below, the data from the assays is represented as the number of counts detected in 600 fields-of-view by the nCounter® Analysis System. The number of counts correlates with the expression level of the target RNA.

[88] Below, exemplary tag and bridge sequences appropriate for detection of corresponding mrRNAs in Table 1 are shown.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1304

Table 2: Exemplary Tag Sequences

miRNA	Tag Sequence	SEQ ID NO:
hsa-let-7a	GGGCTTGACATCGGCGACACAAAATGCTTCTAACTCGCTGTGAAT	11
hsa-let-7b	GGGATATATTCCTTTTTTTTGCAATTAAACTGCCCAGGCGATCTGTTG	12
hsa-let-7c	GCTGCGTAGTTTATCTGCATCACTCGTACTGAAATGCTCACA	13
hsa-let-7d	GGAGTAGTTTGTCCTTCTGGAATTTCTTCCTTTGATTTTGCCATTTT	14
hsa-let-7e	GTCGAAGTCCTTCGAGTGCATGAGCTGTCTTTCACATGATACATCG	15
hsa-let-7f	GGTGGGGCTTGTCGACTGATAGTAACTGTGGTTCGAGTTATGCG	16
hsa-let-7g	GACGGTCCTAGAAGTCAAAAAGCTGCTTGATCGAAAAAAAGCAG	17
hsa-let-7i	GGTTTTAATAGCGCGAGAACAAGTGAGACGTGTATAGTTGCCATGCTG	18
hsa-miR-1	GCCGTGACCCAACTCTATTCCCATTGCCTCCAAACCAGCCCTTG	19
hsa-miR-100	GAAAAATAAAAAGTAAATTGGGCAATACCACCAAAATCGTTCTTTATGGGGT	20

Table 3: Exemplary Bridge Sequences

miRNA	Bridge Sequence	SEQ ID NO:
hsa-let-7a	TGTCAAGCCCAACTATACAACCTACTAC	21
hsa-let-7b	AAGGAATATATCCCAACCACACAACCTAC	22
hsa-let-7c	ACTACGCAGCAACCATACAACCTACT	23
hsa-let-7d	GACAAACTACTCCAACTATGCAACCTACT	24
hsa-let-7e	AGGACTTCGACAACTATACAACCTCC	25
hsa-let-7f	AGCCCCACCAACTATACAATCTACTACC	26
hsa-let-7g	CTAGGACCGTCAACTGTACAAACTACTACC	27
hsa-let-7i	CGCTATTAAAACCAACAGCACAAACTACTAC	28
hsa-miR-1	GGTCACGGCATACATACTTCTTTACATT	29
hsa-miR-100	TTTACTTTTTATTTTTCCACAAGTTCGGATCT	30

Bridge and Tag Design

[89] The addition of a specific nucleic acid tag to the small RNA target serves two main purposes: 1) it adds a robust, unique identifier to each RNA species, making it easier to discriminate between closely-related RNA targets in a hybridization-based assay; and 2) it allows for the manipulation and equalization of the functional T_m's of the targets so that hybridization assays for different RNA targets can be multiplexed at the same temperature.

[90] To avoid cross-hybridization, which would lead to false positive signals, the tag sequence should be unrelated to any other sequence in the transcriptome of the organism from which the small RNA target sequences are derived. Optimally, the tag sequence will be unrelated to the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1305

transcriptome of any known organism, so that the same tags can be used to generate an nCounter® Small RNA Preparation Assay Kit for any organism. Cross-hybridization avoidance was achieved by using novel sequences from the External RNA Controls Consortium (ERCC) database as a starting point for generating the tag sequences.

[91] For ease of synthesis, the nucleic acid tags were designed as DNA oligonucleotides, which were covalently ligated onto the 3' end of each RNA, forming a chimeric RNA-DNA target for detection. The detection assay, based on NanoString's nCounter® Analysis System, relies on the hybridization of two adjacent probes, each with a T_m of 76°C to 84°C, to the target. The final RNA-DNA chimeric target was designed to comprise two regions, a reporter attachment region and a capture attachment region, each with a T_m of approximately 76°C to 84°C.

[92] The ligation of the tags must be highly specific, both individually and in a multiplexed setting. This specificity is achieved through the individual design of DNA oligonucleotide bridges which anneal in a sequence-specific manner to the small RNA at one end and the specific tag at the other end (Figure 2), bringing the RNA and its assigned tag together in close proximity, allowing for a ligation event to form a covalent bond between the two. The ability to multiplex the ligations of hundreds of different species of small RNAs to their respective tags was achieved by careful design of the bridge and tag sequences based on calculations of the T_m's of the bridge/RNA/tag interactions so that all of the bridges anneal specifically to their target RNAs and tags within a narrow temperature range, with calculated T_m's of between 44°C and 53°C. An example of these trimmed sequences is shown in Table 4.

[93] All DNA/DNA T_m calculations were done using the methods described in Allawi H.T., SantaLucia J. (1997) Thermodynamics and NMR of internal G-T mismatches in DNA. *Biochemistry* 36: 10581-10594 and SantaLucia J. Jr, Allawi H.T., Seneviratne P.A. (1996). Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* 35: 3555-3562.

[94] All RNA/DNA T_m calculations were done using the methods described in Sugimoto N., Nakano S., Katoh M., Matsumura A., Nakamuta H., Ohmichi T., Yoneyama M., Sasaki M. (1995). Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes, *Biochemistry* 34: 11211-11216 and SantaLucia J. Jr, Allawi H.T., Seneviratne P.A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability, *Biochemistry* 35:

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1306

3555-3562). Note that the adjustments for the Na^+ concentration in the DNA/RNA calculations were done using the DNA duplex coefficients in SantaLucia *et al.*

Tag and Bridge Design

[95] A starting pool of sequence was generated by selecting sequences from the External RNA Controls Consortium (ERCC) database of sequences (publicly available at http://ukpmc.ac.uk/articlerender.cgi?artid=486406).

[96] These sequences were then processed by dividing them into 35-base regions and calculating their predicted T_m's. A 'reporter attachment pool' was created of 35-base sequences with a DNA/DNA T_m between 76°C and 84°C. A second 'general pool' was created from the remaining sequences of random lengths. For each short RNA sequence (shown as a dotted line in Figures 2 and 3), a 35-base sequence was arbitrarily selected from the reporter attachment pool and assigned as the reporter attachment region (Figures 2-3, fragment A). The reverse complement of the small RNA, the RNA-RC sequence, was generated, and the predicted RNA/DNA T_m of a duplex in which the RNA-RC was the DNA strand was calculated. Based on the outcome of this calculation, the tag and bridge design for each specific small RNA sequence was assigned to one of two design strategies, Type 1 or Type 2.

Type 1 design.

[97] Type 1 tag and bridge design is pictured in Figure 2. If the calculated RNA/DNA T_m of the RNA-reverse complement DNA duplex (RNA-RC DNA duplex) was between 76°C and 84°C, then fragment A in Figure 2 defined the full-length tag and the short RNA sequence (for example the miRNA sequence) defined the capture attachment region. The bridge sequence (Figure 2, fragments C + D) was subsequently calculated in 2 parts: the small RNA bridge segment (C) and the tag bridge segment (D). The small RNA bridge segment was generated by taking the RNA-RC DNA sequence and identifying the fragment of sequence at the 5'-most end with a calculated RNA/DNA T_m of between 44°C and 53°C. The tag bridge segment was generated by taking the reverse complement of the reporter attachment region (A) and identifying the fragment of sequence at the 3'-most end with a calculated DNA/DNA T_m of between 44°C and 53°C. The full bridge sequence was created by joining the 5' end of the small RNA bridge segment to the 3' end of the tag bridge segment (C+D).

Type 2 design

[98] Type 2 tag and bridge design is pictured in Figure 2. If the calculated T_m of the RNA-RC

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1307

DNA duplex was below 76°C, then an additional segment of DNA sequence of random size was selected from the general pool sequences, added to the 5' end of the RNA-RC, and the predicted DNA/DNA T_m of this combined sequence was calculated. The actual duplex this portion forms in the final hybridization assay is a DNA/RNA-DNA-chimera duplex, but the calculation was simplified to reflect a DNA/DNA interaction. This process was repeated until a sequence combination with a predicted DNA/DNA T_m between 76°C and 84°C was found. This additional segment of sequence is shown as segment B in Figure 3. The full-length tag was then created by joining B to the 5' end of A (Figure 3). The small RNA sequence combined segment B defined the capture attachment region.

[99] The bridge sequence (C+D, Figure 3) was calculated in 2 parts: the small RNA bridge segment (C) and the tag bridge segment (D, Figure 3). The small RNA bridge segment was generated in an identical manner as that described in the Type 1 design above. The tag bridge segment was generated by taking the reverse complement of the Type 2 tag sequence (B+A, Figure 3) and identifying the fragment of sequence at the 3'-most end with a calculated DNA/DNA T_m of between 44°C and 53°C. The full bridge sequence was created by joining the 5' end of the small RNA bridge segment to the 3' end of the tag bridge segment (C+D).

Table 4: Trimming Sequences

miRNA Name	Trimmed Sequence	Length	Tm (RNA:DNA)
let-7a	GUAGUAGGUUGUAUAGUU (SEQ ID NO: 31)	18	48.6753
let-7b	GUAGGUUGUGUGGUU (SEQ ID NO: 32)	15	49.7397
let-7c	AGUAGGUUGUAUGGUU (SEQ ID NO: 33)	16	50.4865
let-7d	AGUAGGUUGCAUAGUU (SEQ ID NO: 34)	16	50.5324
let-7e	GGAGGUUGUAUAGUU (SEQ ID NO: 35)	15	48.3192

Alien DNA Sequences

[100] The term "alien DNA sequence" describes a randomized DNA sequence that has minimal or negligible cross-reactivity with any known genome. The homology and/or identity of a candidate alien DNA sequence is compared to all known sequences in a particular species or all

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1308

species by, for example, using the NBLAST and BLAST programs of Altschul, *et al.* (J. Mol. Biol. 215:403-10 (1990)). Alien DNA sequences do not encode genes or regulatory elements thereof. Rather, the alien DNA sequences have non-coding functions within the compositions, methods, and kits of the invention.

[101] Alien DNA sequences are used within the compositions, methods, and kits of the invention to minimize or prevent interactions (*e.g.* hybridization) between molecules that contain alien DNA sequences and molecules that are isolated from a target sample, genome, cell, animal, or extract thereof (*e.g.* target RNA molecules, non-corresponding bridges, non-corresponding tags, non-corresponding reporter molecules). Target RNA molecules and the corresponding bridges, and tags share either partial or complete sequence complementarity, whereas target RNA molecules do not share either partial or complete sequence complementarity with non-corresponding bridges or tags. Reporter molecules and the corresponding tags share either partial or complete sequence complementarity, whereas reporter molecules do not share either partial or complete sequence complementarity with non-corresponding tags.

[102] Tag molecules, reporter molecules, and bridge molecules contain alien DNA sequences. The alien DNA molecules contained within a tag and a corresponding bridge specifically hybridize to each other, and optimally, to no other molecule present in the composition during the denaturing, annealing, ligation, and/or purification reaction. Alternatively, the alien DNA molecules contained within a tag and a corresponding bridge specifically hybridize to each other, and preferably, have a lower than 15% homology across 35-50 bases with a maximum continuous homology of 15 bases or less to the other molecules present in the composition during the denaturing, annealing, ligation, and/or purification reaction. A reporter molecule containing an alien DNA sequence may specifically hybridize to the corresponding tagged miRNA molecule following the purification reaction, and optimally, to no other molecule present in the composition. Alternatively, a reporter molecule containing an alien DNA sequence may specifically hybridize to the corresponding tagged miRNA molecule following the purification reaction, and preferably, have a lower than 15% homology across 35-50 bases with a maximum continuous homology of 15 bases or less to the other molecules present in the composition.

[103] Exemplary alien DNA sequences encompassed by the invention have no greater than 85% identity across 35-50 bases with a maximum continuous homology of 15 bases or less to any known genomic DNA or RNA sequence. Moreover, the alien DNA has no greater than 85%

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1309

identity across 35-50 bases with a maximum continuous homology of 15 bases or less to any known DNA or RNA sequence from or transcribed by a genome from which a target RNA molecule is transcribed. Target RNA molecules are from or transcribed by the genomes of any species, including, but not limited to, animals, plants, bacteria, fungi and viruses. Preferred species from which target RNA molecules are derived include mammals, and most preferably include humans, primates, and mice.

[104] Alien DNA sequences have a lower than 15% homology across 35-50 bases with a maximum continuous homology of 15 bases or less to all known DNA and RNA molecules. Specifically, alien DNA sequences have a homology of less than 15% across 35-50 bases with a maximum continuous homology of 15 bases or less to all the DNA and RNA molecules in a composition or kit. Specifically, the alien DNA sequence has a lower than 15%, 10%, 5%, 2%, 1% or any percentage point in between, homology across 35-50 bases, with a maximum continuous homology of 15 bases or less of DNA and RNA molecules in a composition or kit.

[105] Alien DNA sequences may hybridize to a portion of a bridge that contains DNA, or specifically, alien DNA. Specifically, the alien DNA sequence have a lower than 15%, 10%, 5%, 2%, 1% or any percentage point in between, homology across 35-50 bases with a maximum continuous homology of 15 bases or less with any DNA and RNA molecules in the composition other than the third DNA sequence of the bridge.

Nucleic Acid Molecules

[106] The invention provides bridges and tags, *e.g.* isolated and purified nucleic acid molecules that bind small RNA molecules, including miRNA molecules. Exemplary tag molecules include, but are not limited to, those deoxyribonucleic acid molecules of SEQ ID NOs: 11-20. Exemplary bridge molecules include, but are not limited to, those deoxyribonucleic acid molecules of SEQ ID NOs: 21-30.

[107] In other aspects of the invention, isolated nucleic acid molecules, such as RNA molecules and DNA bridges and tags, bind sequences containing one or more insertions, deletions, inversions, frameshifts, translocations, recombinations, or substitutions.

[108] The isolated nucleic acid molecules of the present invention include single- and double-stranded ribonucleic acid (RNA) and deoxyribonucleic acid (DNA) molecules, as well as all art-recognized analogs, derivatives, or hybrid molecules thereof. Isolated nucleic acid molecules of the present invention also include reagents for synthesizing RNA molecules, bridges, and tags, such as

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1310

isolated full-length genes, transcripts, cDNA molecules, primers, vectors, plasmids, endogenous or naturally-occurring RNA molecules, bridges, tags, and fragments thereof.

[109] Isolated nucleic acid molecules of the invention are engineered to bind RNA sequences, DNA, or "alien" DNA sequences that are distinct from most other nucleic acids present in the natural source of the nucleic acid molecule. Moreover, an "isolated" nucleic acid molecule can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. A nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered "isolated". Nucleic acid molecules present in nonhuman transgenic animals, which do not naturally occur in the animal, are also considered "isolated". For example, recombinant nucleic acid molecules contained in a vector are considered "isolated". Further examples of "isolated" nucleic acid molecules include recombinant DNA or RNA molecules maintained in heterologous host cells, and purified (partially or substantially) DNA or RNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated nucleic acid molecules of the present invention. Moreover, isolated RNA molecules include, but are not limited to, transfer RNA (tRNA), small nucleolar RNA (snoRNA), microRNA (miRNA), a short interfering RNA (siRNA), short hairpin RNA (shRNA), repeat associated small interfering RNA (rasiRNA), and piwi-interacting RNA (piRNA). Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

[110] Generally, an isolated nucleic acid molecule comprises one or more sequences engineered to bind to a RNA molecule, bridge, or tag, with flanking nucleotide sequences on either side of the target RNA, bridge, or tag sequence. A flanking sequence can include nucleotide residues that are naturally associated with the targeted RNA, bridge or tag and/or heterologous nucleotide sequences. Preferably the flanking sequence is up to about 500, 300, 100, 60, 50, 30, 25, 20, 15, 10, 8, or 4 nucleotides (or any other length in-between) on either side of the target RNA, bridge, or tag sequence, or as long as the full-length gene, entire coding, or non-coding sequence (or any portion thereof, such as, an exon, intron, or a 5' or 3' untranslated region).

[111] As used herein, the term "fragment" is meant to describe an isolated nucleic acid molecule that is shorter the isolated nucleic acid molecule from which it is derived. Fragments of isolated

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1311

nucleic acid molecules of the invention can contain, consist of, or comprise any part of the isolated nucleic acid molecule from which it is derived. A fragment typically comprises a contiguous nucleotide sequence at least about 8 or more nucleotides, more preferably at least about 10 or more nucleotides, and even more preferably at least about 15 or more nucleotides. Further, a fragment could comprise at least about 8, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 250 or 500 (or any other number in-between) nucleotides in length. The length of the fragment will be based on its intended use. A labeled probe can then be used, for example, to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the region of interest. Further, primers and probes can be used in amplification reactions, such as for purposes of assaying RNA or miRNA genes, cross-reactivity of RNA molecules with bridges or tags, homology of RNA molecules with bridge or tag sequences, identity RNA molecules with bridge or tag sequences, or for cloning specific regions of a RNA molecule or gene, bridge molecule, or tag molecule.

[112] An isolated nucleic acid molecule of the present invention further encompasses a RNA molecule, bridge or tag that is the product of any one of a variety of nucleic acid amplification methods, which are used to increase the copy numbers of a polynucleotide of interest in a nucleic acid sample. Such amplification methods are well known in the art, and they include but are not limited to, polymerase chain reaction (PCR) (U.S. Pat. Nos. 4,683,195; and 4,683,202; PCR Technology: Principles and Applications for DNA Amplification, ed. H. A. Erlich, Freeman Press, NY, N.Y., 1992), ligase chain reaction (LCR) (Wu and Wallace, Genomics 4:560, 1989; Landegren *et al.*, Science 241:1077, 1988), strand displacement amplification (SDA) (U.S. Pat. Nos. 5,270,184; and 5,422,252), transcription-mediated amplification (TMA) (U.S. Pat. No. 5,399,491), linked linear amplification (LLA) (U.S. Pat. No. 6,027,923), and the like, and isothermal amplification methods such as nucleic acid sequence based amplification (NASBA), and self-sustained sequence replication (Guatelli *et al.*, Proc. Natl. Acad. Sci. USA 87: 1874, 1990). Based on such methodologies, a person skilled in the art can readily design primers in any suitable regions 5' and 3' to a RNA molecule or gene disclosed herein. Moreover, amplification methods include synthesizing *in silico* a RNA molecule, bridge, or tag described herein.

[113] As used herein, an "amplified polynucleotide" of the invention is an isolated nucleic acid molecule in which the number of molecules has been increased at least two-fold by any

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1312

nucleic acid amplification method performed *in vitro* as compared to its starting amount in a test sample. In other preferred embodiments, an amplified polynucleotide is the result of at least ten-fold, fifty-fold, one-hundred fold, one-thousand fold, or even ten-thousand fold increase as compared to its starting amount in a test sample. Generally, an amplified polynucleotide is at least about 10 nucleotides in length. More typically, an amplified polynucleotide is at least about 15 nucleotides in length. In a preferred embodiment of the invention, an amplified polynucleotide is at least about 20-25 nucleotides in length. In a more preferred embodiment of the invention, an amplified polynucleotide is at least about 20, 25, 30, 35, 40, 45, 50, or 100 nucleotides, or any value in between, in length. In yet another preferred embodiment of the invention, an amplified polynucleotide is at least about 100, 200, or 300 nucleotides, or any value in between in length. While the total length of an amplified polynucleotide of the invention can be as long as an exon, an intron, a 5' UTR, a 3' UTR, or the entire gene where the RNA molecule of interest resides, an amplified product is typically no greater than about 1,000 nucleotides in length (although certain amplification methods may generate amplified products greater than 1000 nucleotides in length).

[114] In one embodiment, the invention provides nucleic acid molecules that comprise, consist essentially of, or consist of, for example, the nucleotide sequences of SEQ ID NOs: 1-35. The techniques described herein may be used to provide an unlimited number and variation of nucleic acid molecules to be used in the assay to enable detection. A nucleic acid molecule comprises a nucleotide sequence when the nucleotide sequence is at least part of the final nucleotide sequence of the nucleic acid molecule. In such a fashion, the nucleic acid molecule can be only the nucleotide sequence or have additional nucleotide residues, such as residues that are naturally associated with it or heterologous nucleotide sequences. Such a nucleic acid molecule can have one to a few additional nucleotides or can comprise many more additional nucleotides. A nucleic acid molecule consists essentially of a nucleotide sequence when the final nucleic acid molecule contains the nucleic acid sequence and only regulatory or maintenance sequences. Regulatory and maintenance sequences are those sequences that allow the nucleotide sequence to be expressed and/or translated. A nucleic acid molecule comprises a nucleotide sequence when the final nucleotide sequence of the nucleic acid molecule contains only that nucleic acid sequence. A brief description of how various types of these nucleic acid molecules can be readily made and isolated is provided below, and such techniques are well known to those

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1313

of ordinary skill in the art (Sambrook and Russell, 2000, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, NY).

[115] Isolated nucleic acid molecules can be in the form of RNA, such as mRNA or miRNA, or in the form DNA, including cDNA and genomic DNA, which may be obtained, for example, by molecular cloning or produced by chemical synthetic techniques or by a combination thereof (Sambrook and Russell, 2000, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, NY). Furthermore, isolated nucleic acid molecules can also be partially or completely in the form of one or more types of nucleic acid analogs, such as peptide nucleic acid (PNA) (U.S. Pat. Nos. 5,539,082; 5,527,675; 5,623,049; 5,714,331). The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the complementary non-coding strand (anti-sense strand). DNA, RNA, or PNA segments can be assembled, for example, from fragments of the human genome or any genome including alien sequences (in the case of DNA or RNA) or single nucleotides, short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic nucleic acid molecule. Nucleic acid molecules can be readily synthesized using the sequences provided herein as a reference; oligonucleotide and PNA oligomer synthesis techniques are well known in the art (see, *e.g.*, Corey, "Peptide nucleic acids: expanding the scope of nucleic acid recognition", Trends Biotechnol. 1997 June; 15(6):224-9, and Hyrup *et al.*, "Peptide nucleic acids (PNA): synthesis, properties and potential applications", Bioorg Med Chem. 1996 January; 4(1):5-23). Furthermore, large-scale automated oligonucleotide/PNA synthesis (including synthesis on an array or bead surface or other solid support) can readily be accomplished using commercially available nucleic acid synthesizers, such as the Applied Biosystems (Foster City, Calif.) 3900 High-Throughput DNA Synthesizer or Expedite 8909 Nucleic Acid Synthesis System, and the sequence information provided herein.

[116] The present invention encompasses nucleic acid analogs that contain modified, synthetic, or non-naturally occurring nucleotides or structural elements or other alternative/modified nucleic acid chemistries known in the art. Such nucleic acid analogs are useful, for example, as detection reagents (*e.g.*, primers/probes). Furthermore, kits/systems (such as beads, arrays, etc.) that include these analogs are also encompassed by the present invention. For example, PNA oligomers that are based on the polymorphic sequences of the present invention are specifically contemplated. PNA oligomers are analogs of DNA in which the phosphate backbone is replaced with a

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1314

peptide-like backbone (Lagriffoul *et al.*, Bioorganic & Medicinal Chemistry Letters, 4: 1081-1082 (1994), Petersen *et al.*, Bioorganic & Medicinal Chemistry Letters, 6: 793-796 (1996), Kumar *et al.*, Organic Letters 3(9): 1269-1272 (2001), W096/04000). PNA hybridizes to complementary RNA or DNA with higher affinity and specificity than conventional oligonucleotides and oligonucleotide analogs. The properties of PNA enable novel molecular biology and biochemistry applications unachievable with traditional oligonucleotides and peptides.

[117] The term "isolated nucleic acid molecule" is not limited to molecules containing only naturally-occurring RNA or DNA, but also encompasses chemically-modified nucleotides and non-nucleotides. In certain embodiments, the isolated nucleic acid molecules lack 2'-hydroxy (2'-OH) containing nucleotides. Non-limiting examples of chemical modifications that are made in a RNA molecule, bridge, or tag, include without limitation phosphorothioate internucleotide linkages, 2'- deoxyribonucleotides, 2'-0-methyl ribonucleotides, 2'-deoxy-2'-fluoro ribonucleotides, "universal base" nucleotides, "acyclic" nucleotides, 5-C-methyl nucleotides, and terminal glyceryl and/or inverted deoxy abasic residue incorporation. These chemical modifications, when used in RNA molecules, bridges or tags, preserve function and increase the stability of these molecules.

[118] Nucleotide and non-nucleotide linkers can be incorporated into RNA molecules, bridges and tags. A non-nucleotide linker may be comprised of an abasic nucleotide, polyether, polyamine, polyamide, peptide, carbohydrate, lipid, polyhydrocarbon, or other polymeric compounds (*e.g.*, polyethylene glycols such as those having from 2 to 100 ethylene glycol units). Specific examples include those described by Seela and Kaiser, Nucleic Acids Res. 18:6353, 1990 and Nucleic Acids Res. 15:3113, 1987; Cload and Schepartz, J. Am. Chem. Soc. 113:6324, 1991; Richardson and Schepartz, J. Am. Chem. Soc. 113:5109, 1991; Ma, *et al.*, Nucleic Acids Res. 21:2585, 1993 and Biochemistry 32:1751, 1993; Durand, *et al.*, Nucleic Acids Res. 18:6353, 1990; McCurdy, *et al.*, Nucleosides & Nucleotides 10:287, 1991; Jschke, *et al.*, Tetrahedron Lett. 34:301, 1993; Ono, *et al.*, Biochemistry 30:9914 (1991); Arnold, *et al.*, International Publication No. WO 89/02439; Usman, *et al.*, International Publication No. WO 95/06731; Dudycz, *et al.*, International Publication No. WO 95/11910 and Ferentz and Verdine, J. Am. Chem. Soc. 113:4000, 1991. A "non-nucleotide" further means any group or compound that can be incorporated into a nucleic acid chain in the place of one or more nucleotide units, including either sugar and/or phosphate substitutions, and allows the remaining bases to exhibit

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1315

their enzymatic activity. The group or compound can be abasic in that it does not contain a commonly recognized nucleotide base, such as adenosine, guanine, cytosine, uracil or thymidine, *e.g.*, at the C1 position of the sugar.

[119] Additional examples of nucleic acid modifications that improve the binding properties and/or stability of a nucleic acid include the use of base analogs such as inosine, intercalators (U.S. Pat. No. 4,835,263) and the minor groove binders (U.S. Pat. No. 5,801,115). Thus, references herein to nucleic acid molecules include PNA oligomers and other nucleic acid analogs. Other examples of nucleic acid analogs and alternative/modified nucleic acid chemistries known in the art are described in Current Protocols in Nucleic Acid Chemistry, John Wiley & Sons, N.Y. (2002). Isolated nucleic acids of the inventions are comprised of base analogs including, but not limited to, any of the known base analogs of DNA and RNA such as, but not limited to 4-acetylcytosine, 8- hydroxy-N-6-methyladenosine, aziridinylcytosine, pseudoisocytosine, 5- (carboxyhydroxylmethyl) uracil, 5-bromouracil, 5-carboxymethylaminomethy1-2- thiouracil, 5-carboxymethylaminomethyluracil, dihydrouracil, inosine, N6- isopentenyladenine, 1-methyladenine, 1-methylpseudouracil, 1-methyl guanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3- methylcytosine, 5-methylcytosine, N6-methyladenine, 7-methylguanine, 5- methylaminomethyluracil, 5-methoxy-aminomethyl-2-thiouracil, beta-Dmannosylqueosine, 5'-methoxycarbonylmethyluracil, 5-methoxyuracil, 2-methylthio-N6- isopentenyladenine, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, oxybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2- thiouracil, 4-thiouracil, 5-methyluracil, N-uracil-5-oxyacetic acid methylester, 2,6-diaminopurine, and 2'-modified analogs such as, but not limited to 0-methyl, amino-, and fluoro-modified analogs.

[120] In certain embodiments, the RNA molecules, bridges and tags of the invention are modified to enhance stability by modification with nuclease resistant groups, *e.g.*, 2'-amino, 2'-C-allyl, 2'-fluoro, 2'-O-methyl, and 2'-H. (For a review see Usman and Cedergren, TIBS 17:34, 1992; Usman, *et al.*, Nucleic Acids Symp. Ser. 31:163, 1994).

[121] Chemically synthesizing nucleic acid molecules with modifications (base, sugar and/or phosphate) prevents their degradation by ribonucleases, which increases the efficiency of the assays described herein. See, *e.g.*, Eckstein, *et al.*, International Publication No. WO 92/07065; Perrault, *et al.*, Nature 344:565, 1990; Pieken, *et al.*, Science 253:314, 1991; Usman and Cedergren,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1316

Trends in Biochem. Sci. 17:334, 1992; Usman, et al, International Publication No. WO 93/15187; and Rossi, *et al.*, International Publication No. WO 91/03162; Sproat, U.S. Pat. No. 5,334,711; Gold, *et al.*, U.S. Pat. No. 6,300,074. All of the above references describe various chemical modifications that are made to the base, phosphate and/or sugar moieties of the isolated nucleic acid molecules described herein.

[122] In one embodiment, the invention provides RNA molecules, bridges, and tags with phosphate backbone modifications comprising one or more phosphorothioate, phosphorodithioate, methylphosphonate, phosphotriester, morpholino, amidate carbamate, carboxymethyl, acetamidate, polyamide, sulfonate, sulfonamide, sulfamate, formacetal, thioformacetal, and/or alkylsilyl, substitutions. For a review of oligonucleotide backbone modifications, see Hunziker and Leumann, "Nucleic Acid Analogues: Synthesis and Properties, in Modern Synthetic Methods," VCH, 331-417, 1995, and Mesmaeker, et al, "Novel Backbone Replacements for Oligonucleotides, in Carbohydrate Modifications in Antisense Research," ACS, 24-39, 1994.

[123] Further variants of the nucleic acid molecules including, but not limited to those identified as SEQ ID NOs: 1-35, such as naturally occurring allelic variants (as well as orthologs and paralogs) or synthetic variants produced by mutagenesis techniques, can be identified and/or produced using methods well known in the art. Such further variants can comprise a nucleotide sequence that shares at least 70-80%, 80-85%, 85-90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with a nucleic acid sequence disclosed as SEQ ID NOs: 1-35 (or a fragment thereof). Thus, the present invention specifically contemplates isolated nucleic acid molecule that have a certain degree of sequence variation compared with the sequences of SEQ ID NOs: 1-35.

[124] The RNA molecules, bridges and tags of the invention are routinely made through techniques such as solid phase synthesis. Equipment for such synthesis is sold by several vendors including, for example, Applied Biosystems, (Foster City, Calif.). Any other means for such synthesis known in the art is additionally or alternatively employed. It is well known to use similar techniques to prepare oligonucleotides such as the phosphorothioates and alkylated derivatives.

[125] Oligonucleotides are synthesized using protocols known in the art, *e.g.*, as described in Caruthers, *et al.*, Methods in Enzymology 211:3-19, 1992; Thompson, *et al.*, International PCT Publication No. WO 99/54459; Wincott, *et al.*, Nucleic Acids Res. 23:2677-2684, 1995; Wincott, *et al.*, Methods Mol. Bio. 74:59, 1997; Brennan, *et al.*, Biotechnol Bioeng. 61:33-45, 1998; and

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1317

Brennan, U.S. Pat. No. 6,001,311. Synthesis of RNA molecules follows general procedures as described, *e.g.*, in Usman, et al, J. Am. Chem. Soc. 109:7845, 1987; Scaringe, *et al.*, Nucleic Acids Res. 18:5433, 1990; and Wincott, *et al.*, Nucleic Acids Res. 23:2677-2684, 1995; Wincott, *et al.*, Methods Mol. Bio. 74:59, 1997.

[126] The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. (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 1, 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). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch algorithm (J. Mol. Biol. (48):444-453 (1970)) which has been incorporated into the GAP program in the GCG software package, using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6.

[127] In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., *et al.*, Nucleic Acids Res. 12(1):387 (1984)), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Myers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4.

[128] The nucleotide and amino acid sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and BLAST programs (version 2.0) of Altschul, *et al.* (J. Mol. Biol. 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score=100, wordlength=12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score=50, wordlength=3 to

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1318

obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (Nucleic Acids Res. 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (*e.g.*, BLAST and NBLAST) can be used. In addition to BLAST, examples of other search and sequence comparison programs used in the art include, but are not limited to, FASTA (Pearson, Methods Mol. Biol. 25, 365-389 (1994)) and KERR (Dufresne *et al.*, Nat Biotechnol 2002 December; 20(12): 1269-71). For further information regarding bioinformatics techniques, see Current Protocols in Bioinformatics, John Wiley & Sons, Inc., N.Y.

Therapeutic Applications

[129] Compositions and methods of the invention are used to detect gene expression in subjects who are at risk of developing an illness or disorder. Moreover, the compositions and methods of the invention are used to detect gene expression in subjects who have been diagnosed with an illness or disorder, and who are in need of a diagnosis or prognosis. The compositions and methods described herein are used to monitor disease progression and the effectivness of therapy on the level of gene expression and regulation. Futhermore, the compositions and methods provided herein are used to screen individuals for their personal risk of developing a disorder as well as their risk of passing a disorder onto future children. Embryonic cells are tested using the compositions and methods of the invention for the presence or absence of disorders.

[130] The invention can be used to determine the risk of developing a particular biological condition, a particular disease, such as a cancer, a genetic disorder, a developmental disorder, a degenerative disorder, a neurological disorder, a stem cell disorder, or other biological condition. Futhermore, the present invention can be used to monitor the progression of a disease or monitor responses to therapy. Specifically, the invention can be used to detect, to monitor progression of, or monitor therapeutic regimens for diseases of the heart, kidney, ureter, bladder, urethra, liver, prostate, heart, blood vessels, bone marrow, skeletal muscle, smooth muscle, various specific regions of the brain (including, but not limited to the amygdala, caudate nucleus, cerebellum, corpus callosum, fetal, hypothalamus, thalamus), spinal cord, peripheral nerves, retina, nose, trachea, lungs, mouth, salivary gland, esophagus, stomach, small intestines, large intestines, hypothalamus, pituitary, thyroid, pancreas, adrenal glands, ovaries, oviducts, uterus, placenta, vagina, mammary glands, testes, seminal vesicles, penis, lymph nodes, thymus, and spleen. The

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1319

present invention can be used to detect, to monitor progression of, or monitor therapeutic regimens for a particular disease, such as a cancer, a genetic disorder, a developmental disorder, a degenerative disorder, a neurological disorder, a stem cell disorder, or other biological condition.

[131] The methods of the invention encompass a variety of subjects, including mammals. In certain embodiments, the mammal is a human, non-human primate, mouse, rat, dog, cat, horse, or cow, but is not limited to these examples. Mammals other than humans are advantageously used as subjects that represent animal models of a particular disorder. The preferred subject is human. The methods of the invention also encompass a variety of subjects, including non-mammals. These include birds, reptiles, plants, bacteria, fungi, protists and viruses. Any sample containing naturally occurring or synthetic small RNA molecules may be used as the subject of the assay.

Cancer

[132] Compositions and methods of the invention are used to identify cells and subjects at risk of developing or those cells and subjects who may have a predisposition for developing cancer. Moreover, the compositions and methods of the invention are used to differentiate cancer cell type, cancer subtype, tumor grade, or cancer stage for the purpose of diagnosing or prognosing a subject at risk of developing cancer or a subject who has developed cancer. The compositions and methods of the invention are further used to monitor to progression of a tumor, cancer, or a treatment regime. Additionally, the compositions and methods of the invention are used to screen individuals for any genetic predisposition to developing cancer.

[133] The term "cancer" includes solid tumors, as well as, hematologic tumors and/or malignancies. A "precancer cell" or "precancerous cell" is a cell manifesting a cell proliferative disorder that is a precancer or a precancerous condition. A "cancer cell" or "cancerous cell" is a cell manifesting a cell proliferative disorder that is a cancer. Any reproducible means of measurement may be used to identify cancer cells or precancerous cells. Cancer cells or precancerous cells can be identified by histological typing or grading of a tissue sample (*e.g.*, a biopsy sample). Cancer cells or precancerous cells can be identified through the use of appropriate molecular markers.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1320

[134] The compositions and methods of the invention are used to further determine cancer severity, as it is characterized by stage, tumor grade, and expression of factors that degrade the extracellular matrix, induce vascularization, inhibit cell adhesion and enable metastasis.

[135] Exemplary cancers include, but are not limited to, adrenocortical carcinoma, AIDS-related cancers, AIDS-related lymphoma, anal cancer, anorectal cancer, cancer of the anal canal, appendix cancer, childhood cerebellar astrocytoma, childhood cerebral astrocytoma, basal cell carcinoma, skin cancer (non-melanoma), biliary cancer, extrahepatic bile duct cancer, intrahepatic bile duct cancer, bladder cancer, uringary bladder cancer, bone and joint cancer, osteosarcoma and malignant fibrous histiocytoma, brain cancer, brain tumor, brain stem glioma, cerebellar astrocytoma, cerebral astrocytoma/malignant glioma, ependymoma, medulloblastoma, supratentorial primitive neuroectodeimal tumors, visual pathway and hypothalamic glioma, breast cancer, bronchial adenomas/carcinoids, carcinoid tumor, gastrointestinal, nervous system cancer, nervous system lymphoma, central nervous system cancer, central nervous system lymphoma, cervical cancer, childhood cancers, chronic lymphocytic leukemia, chronic myelogenous leukemia, chronic myeloproliferative disorders, colon cancer, colorectal cancer, cutaneous T-cell lymphoma, lymphoid neoplasm, mycosis fungoides, Seziary Syndrome, endometrial cancer, esophageal cancer, extracranial germ cell tumor, extragonadal germ cell tumor, extrahepatic bile duct cancer, eye cancer, intraocular melanoma, retinoblastoma, gallbladder cancer, gastric (stomach) cancer, gastrointestinal carcinoid tumor, gastrointestinal stromal tumor (GIST), germ cell tumor, ovarian germ cell tumor, gestational trophoblastic tumor glioma, head and neck cancer, hepatocellular (liver) cancer, Hodgkin lymphoma, hypopharyngeal cancer, intraocular melanoma, ocular cancer, islet cell tumors (endocrine pancreas), Kaposi Sarcoma, kidney cancer, renal cancer, kidney cancer, laryngeal cancer, acute lymphoblastic leukemia, acute myeloid leukemia, chronic lymphocytic leukemia, chronic myelogenous leukemia, hairy cell leukemia, lip and oral cavity cancer, liver cancer, lung cancer, non-small cell lung cancer, small cell lung cancer, AIDS-related lymphoma, non-Hodgkin lymphoma, primary central nervous system lymphoma, Waldenstram macroglobulinemia, medulloblastoma, melanoma, intraocular (eye) melanoma, merkel cell carcinoma, mesothelioma malignant, mesothelioma, metastatic squamous neck cancer, mouth cancer, cancer of the tongue, multiple endocrine neoplasia syndrome, mycosis fungoides, myelodysplastic syndromes, myelodysplastic/ myeloproliferative diseases, chronic myelogenous leukemia, acute

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1321

myeloid leukemia, multiple myeloma, chronic myeloproliferative disorders, nasopharyngeal cancer, neuroblastoma, oral cancer, oral cavity cancer, oropharyngeal cancer, ovarian cancer, ovarian epithelial cancer, ovarian low malignant potential tumor, pancreatic cancer, islet cell pancreatic cancer, paranasal sinus and nasal cavity cancer, parathyroid cancer, penile cancer, pharyngeal cancer, pheochromocytoma, pineoblastoma and supratentorial primitive neuroectodermal tumors, pituitary tumor, plasma cell neoplasm/multiple myeloma, pleuropulmonary blastoma, prostate cancer, rectal cancer, renal pelvis and ureter, transitional cell cancer, retinoblastoma, rhabdomyosarcoma, salivary gland cancer, ewing family of sarcoma tumors, Kaposi Sarcoma, soft tissue sarcoma, uterine cancer, uterine sarcoma, skin cancer (non-melanoma), skin cancer (melanoma), merkel cell skin carcinoma, small intestine cancer, soft tissue sarcoma, squamous cell carcinoma, stomach (gastric) cancer, supratentorial primitive neuroectodermal tumors, testicular cancer, throat cancer, thymoma, thymoma and thymic carcinoma, thyroid cancer, transitional cell cancer of the renal pelvis and ureter and other urinary organs, gestational trophoblastic tumor, urethral cancer, endometrial uterine cancer, uterine sarcoma, uterine corpus cancer, vaginal cancer, vulvar cancer, and Wilm's Tumor.

Developmental and Degenerative Disorders

[136] Compositions and methods of the invention are used to identify cells and subjects at risk of developing a developmental or degenerative disorder or those cells and subjects who may have a predisposition for developing a developmental or degenerative disorder. Moreover, the compositions and methods of the invention are used to differentiate developmental disorders, degenerative disorders, or developmental from degenerative disorders for the purpose of diagnosing or prognosing a subject at risk of presenting or a subject who has been diagnosed with a developemental or degenerative disorder. The compositions and methods of the invention are further used to monitor to progression of a developmental disorder, a degenerative disorder, or a treatment regime. Additionally, the compositions and methods of the invention are used to screen individuals for any genetic predisposition for presenting a developmental or degenerative disorder himself/herself, or for producing a child having a developmental or degenerative disorder.

[137] The term "developmental disorder" includes any disorder that initially presents in an individual during gestation or early postnatal development. Early postnatal development encompasses a period of time from birth to age 18. Although developmental disorders are often

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1322

considered synonymous with metnal disabilities that cause mental, emotional, or cognitive deficits, the term "developmental disorder" is meant to encompass any disorder that presents in either a fetus or a child aged 18 years or less, regardless of the specific signs or symptoms associated with the disorder. Moreover, developmental disorders are typically characterized by an inadequate or malfunctioning development of biological or psychological process. Developmental disorders are also characterized by behavioral traits, family history, brain morphology, or genetic/biomarkers that are present during development and predict or indicate the individual's risk of developing the disease in adulthood (*e.g.* Huntington's Disease, Amyotrophic lateral sclerosis or ALS, and Schizophrenia).

[138] A specific developmental disorder selectively affects one area of development, sparing essentially all other areas of development. Specific developmental diorders affect primarily hearing, vision, speech, or metabolism. However, a pervasive developmental disorder involves delays in the development of many basic skills, most notably the ability to socialize with others, because these conditions affect the child's ability to communicate and to use imagination. Pervasive developmental disorders include, but are not limited to, austism and autism spectrum disorders, Asperger's syndrome, childhood disintegrative disorder, Rett's Syndrome, attention-deficit disorder (ADD), and unspecified but pervasive disorders.

[139] Exemplary developmental disorders also include, but are not limited to, Autism spectrum disorders (ASD), Angelman Syndrome, central auditory processing disorder (CAPD), cerebral palsy, Down Sydrome, expressive language disorder, IsoDendric 15 (abbreviated idic(15)), Lanau-Kleffner Syndrome, neural tube defects, phenylketonuria (PKU), Prader-Willi Syndrome, seizure disorders, epilepsy, Tourette Syndrome, Williams Syndrome, hearing loss, deafness, blindness, vision impairment, jaundice / kernicterus, cluttering (speech disfluency), agnosias (visual, auditory, and somatosensory), anorexia nervosa disorder, acute stress disorder, adjustment disorder, bipolar disorder, body dysmorphic disorder, breathing-related sleep disorders, asthma, brief psychotic episode, bulimia nervosa, schizophrenia, Huntington's Disease (HD), multiple sclerosis (MS), amyotrophic lateral sclerosis (ALS), chronic motor or vocal tic disorder, circadian rhythm sleep disorder, conduct disorder, communication/language disorders, Cornelia de Lange Syndrome, fatal familial insomnia (FFI), Fahr's Syndrome (or idiopathic basal ganglia calcification), migraine, neoplasm (benign and malignant), Lupus erythematosus, autoimmune disorders, diabetes (type I), Wilson's Disease, Bell's Palsy, congenital heart disease,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1323

microcephaly, neonatal encephalitis, hydrocephalis, Parkinson's Disease, narcolepsy, muscular distrophy, Guillain-Barre Syndrome, neurofibromatosis, Von Hippel-Lindau Disease, dyslexia, familial hypercholesterolemia, polycystic kidney disease, hereditary spherocytosis, hereditary breast and ovarian syndrome, marfan syndrome, sickle cell anemia, sickle cell disease, cystic fibrosis, mucopolysaccharidoses, glycogen storage diseases, glactosemia, hemophilia, Androgenetic alopecia, Lebner's hereditary optic neuropathy, autoimmune disease, cleft palate, obesity, Gauchers Disease, Rett Syndrome, ataxia telagiectasia, long QT Syndrome, Alport Syndrome, male pattern baldness, SRY sex determination, achondroplasia, Cockayne syndrome, DiGeorge syndrome, fragile X syndrome, severe combined immunodeficiency, Waardenburg syndrome, Werner syndrome, Zellweger syndrome, adrenoleukodystrophy, glucose galactose malabsorption, hereditary hemochromatosis, Lesch-Nyhan syndrome, maple syrup urine disease, Menkes syndrome, Neimann-Pick syndrome, porphyria, Refsum disease, Tangier disease, Tay-Sachs disease, diastropic dysplasia, Ellis-van Creveld Syndrome (chondroectodermal dysplasia), paroxysmal nocturnal hemoglobinuria, thalassemia, Crohn's disease, Best disease, glaucoma, retinoblastoma, congenital adrenal hyperplasia, autoimmune polyglandular syndrome, multiple endocrine neoplasia, familial Mediterranean fever, immunodeficiency with hyper- IgM, Charcot-Marie-Tooth syndrome, fibrodysplasia ossificans progressive, myotonic dystrophy, essential tremor, Friedrich's Ataxia, spinal muscular atrophy, spinocerebellar ataxia, tuberous sclerosis, alpha-1-antitrypsin deficiency, and Pendred Syndrome.

[140] The term "degenerative disorder" includes any disorder that initially presents in an adult individual. The term adult encompasses a period of time from age 18 to death. Although degenerative disorders are often considered synonymous with mental disabilities that cause mental, emotional, or cognitive deficits, the term "degenerative disorder" is meant to encompass any disorder that presents in an adult aged 18 years or older, regardless of the specific signs or symptoms associated with the disorder. Moreover, degenerative disorders are typically characterized by the deregulation or malfunction of an ordinarly operable biological or psychological process. Degenerative disorders can result from genetic predisposition, environmental factors, or exposure to pathogens such as a virus or prion.

[141] Exemplary degenerative disorders include, but are not limited to, Alzheimer's Disease, dementia, senility, agnosias (visual, auditory, and somatosensory), acute stress disorder, adjustment disorder, bipolar disorder, body dysmorphic disorder, breathing-related sleep

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1324

disorders (sleep apnea), brief psychotic episode, bulimia nervosa, schizophrenia, Huntington's Disease (HD), Parkinson's Disease, multiple sclerosis (MS), amyotrophic lateral sclerosis (ALS), Capgras (delusion) Syndrome, chronic fatique syndrome, circadian rhythm sleep disorder, conduct disorder, communication/language disorders, Creutzfeldt-Jakob Disease (CJD), kuru, Gerstmann–Sträussler–Scheinker syndrome (GSS), fatal familial insomnia (FFI), cyclothymic disorder, acquired immune deficiency syndrome (AIDS), depression, addiction, Cushing's Syndrome (also called hyperadrenocorticism or hypercorticism), neoplasm (benign and malignant), stroke, diabetes (Type II), aneurysm, cardiovascular disease (including heart disease), Meniere's Disease, deafness, blindness, multiple system atrophy, Neiman Pick Disease, artherosclerosis, progressive supranuclear palsy, cancer, Tay-Sachs Disease, keratoconus, macular degeneration, inflammatory bowel disease (IBD), prostatis, male pattern baldness, obesity, paroxysmal nocturnal hemoglobinuria, thalassemia, Crohn's disease, Best disease, glaucoma, Gyrate atrophy of the choroid and retina, Charcot-Marie-Tooth syndrome, fibrodysplasia ossificans progressive, myotonic dystrophy,osteoarthritis, osteoporosis, arthritis, and rheumatoid arthritis.

Neurological Disorders

[142] Compositions and methods of the invention are used to identify cells and subjects at risk of developing a neurological disorder or those cells and subjects who may have a predisposition for developing a neurological disorder. Moreover, the compositions and methods of the invention are used to differentiate neurological disorders for the purpose of diagnosing or prognosing a subject at risk of presenting or a subject who has been diagnosed with a neurological disorder. The compositions and methods of the invention are further used to monitor to progression of a neurological disorder or a treatment regime. Additionally, the compositions and methods of the invention are used to screen individuals for any genetic predisposition for presenting a neurological disorder himself/herself, or for producing a child having a neurological disorder.

[143] The term "neurological disorder" includes any disorder that initially presents within the nervous system of an individual. Neurological disorders present with a variety of signs and symptoms including, but not limited to, psychological, mood, or behaviorial changes; loss or decreased accuity of one or more senses (vision, hearing, touch); increased pain or burning sensations; lack of coordination or balance; loss of memory; loss of control over voluntary or involuntary movement; speech or balance; visual or auditory hallucinations; seizures; headaches;

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 1325