

Provisional Application for Patent Cover Sheet

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c)

Inventor(s)

Inventor 1

Remove

Given Name

Middle Name

Family Name

City

State

Country i

Jian-Bing

Fan

San Diego

CA

US

Inventor 2

Remove

Given Name

Middle Name

Family Name

City

State

Country i

Kun

Zhang

San Diego

CA

US

All Inventors Must Be Listed – Additional Inventor Information blocks may be generated within this form by selecting the **Add** button.

Add

Title of Invention

METHODS AND COMPOSITIONS FOR SINGLE CELL GENOMICS

Attorney Docket Number (if applicable)

IP-0645-PRV-A

Correspondence Address

Direct all correspondence to (select one):

☒ The address corresponding to Customer Number

☐ Firm or Individual Name

Customer Number

29389

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☒ No.

☐ Yes, the invention was made by an agency of the United States Government. The U.S. Government agency name is:

☐ Yes, the invention was under a contract with an agency of the United States Government. The name of the U.S. Government agency and Government contract number are:

Entity Status

Applicant asserts small entity status under 37 CFR 1.27 or applicant certifies micro entity status under 37 CFR 1.29

- ☐ Applicant asserts small entity status under 37 CFR 1.27
- ☐ Applicant certifies micro entity status under 37 CFR 1.29. Applicant must attach form PTO/SB/15A or B or equivalent.
- ☒ No

Warning

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

Signature

Please see 37 CFR 1.4(d) for the form of the signature.

Signature	/brentcmoore/			Date (YYYY-MM-DD)	2014-06-19
First Name	Brent C.	Last Name	Moore	Registration Number (If appropriate)	55461

This collection of information is required by 37 CFR 1.51. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 8 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **This form can only be used when in conjunction with EFS-Web. If this form is mailed to the USPTO, it may cause delays in handling the provisional application.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that : (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

METHODS AND COMPOSITIONS FOR SINGLE CELL GENOMICS

BACKGROUND

Historically, many biological questions were addressed by gene expression
5 analysis of many cells. One of the shortcomings of this strategy is that gene expression
heterogeneity, which is present even among genetically identical cells that are grown in
the same culture or are found at the same developmental point, is obscured. As a result,
the gene expression profile of individual cells and the cell-to-cell variability will be lost,
as they become part of the average, and single cell level analyses are required to provide
10 information about these aspects.

A general strategy in existing single-cell transcriptome analysis methods is to
amplify the limited number of RNA molecules from a cell and quantify the transcript
abundance by hybridization or sequencing. A key challenge is that all RNA species in
the target cells need to be amplified specifically and uniformly to a large quantity with
15 little loss. There remains a great need for robust methods that are inherently scalable
and automatable, and that, most importantly, can provide accurate quantitative
information of all RNA species regardless of the length and presence or absence of 3'
poly-A tails.

Another layer of information in single cell expression analysis is the three
20 dimensional map of transcriptional activities in single cells. In the past, after being
isolated, tissue samples were trypsinized and the resulting cell suspensions were used to
examine individual cells, but during this procedure, the spatial relationship among cells
from the original tissue sample was lost. Mapping the specific individual cells back to
the three-dimensional organ or tissue is challenging, but this is the future direction
25 which single cell gene expression profiling will develop. It is fundamental for
understanding cellular functions that shape development and disease.

BRIEF SUMMARY

Presented herein are methods and compositions for obtaining sequence
30 information from one or more individual cells. The methods are useful for obtaining

sequence information for a single nucleotide sequence, and for multiplex generation of sequence information from an individual cell. In some embodiments presented herein, the methods comprise first amplifying one or more nucleic acid molecules *in situ* within a cell, and subsequently hybridizing a barcode probe to the amplified nucleic acid molecules. The identity of the amplified nucleic acid molecules is then determined by decoding of the barcode probes. In other embodiments presented herein, the methods comprise first hybridizing one or more different barcode probes to one or more nucleic acid molecule *in situ* within a cell, and subsequently amplifying the barcode probe, thereby generating an amplified barcode probe. The identity of the nucleic acid molecules is then determined by decoding of the amplified barcode probes.

Accordingly, presented herein is a method of obtaining sequence information from an individual cell comprising: amplifying a nucleic acid molecule *in situ* within a cell, thereby generating an amplified nucleic acid molecule; annealing *in situ* a barcode probe comprising at least a first decoder sequence and a second decoder sequence to the amplified nucleic acid molecule; hybridizing a first labeled decoder probe to the first decoder sequence; detecting a signal indicative of the hybridizing; hybridizing a second labeled decoder probe to the second decoder sequence; and detecting a signal indicative of the hybridizing, thereby generating sequence information regarding the amplified nucleic acid molecule.

Also presented herein is a method of obtaining sequence information regarding a plurality of nucleic acid molecules from an individual cell comprising: amplifying a plurality of nucleic acid molecules *in situ* within a cell, thereby generating a set of amplified nucleic acid molecules; for each of a plurality of amplified nucleic acid molecules in the set, annealing *in situ* a barcode probe comprising at least a first decoder sequence and a second decoder sequence to the amplified nucleic acid molecule; and hybridizing a first labeled decoder probe to the first decoder sequence; detecting a plurality of signals indicative of the hybridizing; hybridizing a second labeled decoder probe to the second decoder sequence; and detecting a plurality of signals indicative of the hybridizing, thereby generating sequence information regarding the plurality of amplified nucleic acid molecules.

Also presented herein is a method of obtaining sequence information from an individual cell comprising: annealing a barcode probe specific for a disease-causing mutation to a nucleic acid molecule *in situ* within a cell, the barcode probe comprising at least a first decoder sequence and a second decoder sequence; amplifying the barcode probe *in situ*, thereby generating an amplified barcode probe; hybridizing a first labeled decoder probe to the first decoder sequence; detecting a signal indicative of the hybridizing; hybridizing a second labeled decoder probe to the second decoder sequence; and detecting a signal indicative of the hybridizing, thereby generating sequence information regarding the amplified nucleic acid molecule.

Also presented herein is a method of obtaining sequence information from a plurality of individual cells in parallel. The cells can be a pool of cell populations from blood or any liquid biospecimen. The cells can also be many cells from a tissue section in which the original spatial relationships of the cells are retained.

In some aspects of the above embodiments, the first labeled decoder probe is removed prior to hybridizing the second labeled decoder probe. In some aspects, the second labeled decoder probe is hybridized prior to the first labeled decoder probe.

In certain aspects, detecting can comprise sequencing at least the first decoder sequence *in situ*. Detecting can also further comprise sequencing at least the second decoder sequence *in situ*.

In certain aspects, amplifying comprises rolling circle amplification. In certain aspects, the barcode probe can comprise a rolling circle amplification probe, such as a padlock probe. In certain aspects, the barcode probe comprises an oligonucleotide ligation assay probe.

In certain aspects, the amplified nucleic acid molecule or molecules are immobilized within the one or more cells.

In certain aspects, the amplified nucleic acid molecule is amplified from an RNA template, for example by cDNA synthesis using the RNA as a template. In certain aspects, the amplified nucleic acid molecule comprises DNA.

In certain aspects, the nucleic acid molecule comprises a disease-causing mutation, allele or common variant. In certain such aspects, the methods can comprise

hybridizing a barcode probe which is correlated with the disease-causing mutation, allele or common variant. In some aspects, the method can comprise ligation of a mutation-specific padlock probe only when the disease-causing mutation, allele or common variant is present in the nucleic acid molecule. The disease-causing mutation, allele or common variant can be, for example, a base substitution, an insertion mutation, a deletion mutation, a gene amplification, a gene deletion, a gene fusion event, a gene inversion invent and the like.

In some aspects, detecting can further comprise evaluating a signal to determine the presence, quantity, frequency, and/or spatial distribution of a disease-causing mutation, allele or common variant within a tissue.

The details of one or more embodiments are set forth in the accompanying drawings and the description below. Other features, objects, and advantages will be apparent from the description and drawings, and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 describes one embodiment of padlock capture and hybridization decoding of rolonies. (A) Molecular barcodes are attached to rolonies in a gene specific manner using padlock probes. (B) After padlock capture, each rolony will have a unique barcode in thousands of copies. These barcodes can be accurately decoded by hybridization similar to the decoding of Illumina bead array, using a combinatorial chemistry approach (Gunderson et al. Genome Res. 14(5):870-877 (2004)). (C) Schematic of padlock probes. In the example shown, each probe has two gene-specific (and/or allele-specific) capturing arms H1/H2. The region in the middle contains six slots for hybridization barcodes. In an alternative embodiment, the padlock probe can be replaced with two independent probes, for example, two oligo ligation probes - one contains the barcode sequences and upstream target-specific sequence and one contains downstream target-specific sequences (and with or without barcode sequences).

DETAILED DESCRIPTION

Existing whole transcriptome amplification methods involve lengthy processing with a large number of experimental steps which do not allow for automation and implementation at a production scale. Thus, there remains a great need for a robust
5 method that is inherently scalable and automatable, and that, importantly, can provide accurate quantitative information of all RNA species regardless of the length and presence or absence of 3' poly-A tails.

Spatial information is also desirable when interpreting transcriptional heterogeneity of individual cells in a structured tissue. However, RNA sequencing is
10 typically performed on isolated cells, such that spatial information is lost. One existing approach to link transcriptome information from isolated cells to their context is to first acquire the spatial information as well as other "phenotypes" of individual cells within the intact tissue using fluorescent microscopy, and then extract specific cells (or cellular content) for subsequent transcriptional profiling. Thus, there is a need for methods
15 which provide high-resolution mapping of a large number of transcripts *in situ*.

Accordingly, presented herein are methods and compositions for obtaining sequence information from an individual cell. The methods are useful for obtaining sequence information at the level of individual nucleic acid molecules. The methods are especially advantageous for multiplex generation of sequence information from an
20 individual cell. As such, sequence information can be obtained for several individual nucleic acid molecules in an individual cell. In some embodiments presented herein, the methods comprise first amplifying one or more nucleic acid molecules *in situ* within a cell, and subsequently hybridizing a barcode probe to the amplified nucleic acid molecules. The identity of the amplified nucleic acid molecules is then determined by
25 decoding of the barcode probes. In other embodiments presented herein, the methods comprise first hybridizing one or more different barcode probes to one or more nucleic acid molecule *in situ* within a cell, and subsequently amplifying the barcode probe, thereby generating an amplified barcode probe. The identity of the nucleic acid molecules is then determined by decoding of the amplified barcode probes.

The methods and compositions set forth herein will be exemplified for detection of nucleic acids in a single cell. However, the methods can be extended to the simultaneous detection of nucleic acids in several individual cells. Parallel detection can be carried out such that several individual cells are resolved with regard to nucleic acid sequences that are present. In some embodiments the spatial location of nucleic acids can be simultaneously determined for several individual cells in a tissue or other cell-containing sample. It will be understood that particular embodiments of the methods set forth herein have broader use and need not be used to resolve individual cells. Rather the methods can be used to detect nucleic acids in a collection of several cells where the cells are not individually resolved with respect to which of the cells do or do not contain a particular sequence present in the collection of cells.

Barcode Probes for Detection of Amplified Nucleic Acid Molecules

The methods and compositions presented herein relate to obtaining sequence information from an individual cell. In some embodiments of the methods provided herein, the methods comprise first amplifying one or more nucleic acid molecules *in situ* within a cell, and subsequently hybridizing a barcode probe to the amplified nucleic acid molecules. The identity of the amplified nucleic acid molecules is then determined by decoding of the barcode probes.

Thus, in certain embodiments, the methods can comprise amplifying a nucleic acid molecule *in situ* within a cell, thereby generating an amplified nucleic acid molecule. As used herein, the term *in situ* refers to analysis methods which are performed within a cell or tissue. In some embodiments, an *in situ* analysis method is performed in a viable cell which is capable of performing one or more cellular functions typically associated with a living cell, such as respiration, cellular transport, transcription, translation and the like. In some embodiments, an *in situ* analysis method is performed in a cell that is not viable, but retains the structure and organization of a viable cell. For example, a cell which has been permeabilized and/or fixed is not considered viable, even though subcellular organelles and molecules such as proteins and nucleic acids remain intact and substantially in the same cellular location compared

to a viable cell. Thus, analysis methods such as amplification, annealing, hybridizing, decoding and the like which are performed *in situ* are performed within the structure of a viable or non-viable cell or tissue. In some embodiments, the individual cell has been isolated from its original tissue, for example, using laser capture microdissection or any other suitable isolation technique as known in the art. In other embodiments, the individual cell remains within the context of its original surrounding tissue. For example, the individual cell can be part of a tissue slice from frozen or paraffin-embedded tissue. Thus, in some embodiments, the methods presented herein can be performed on a plurality of individual cells simultaneously.

Barcode Probes

The methods provided herein further comprise annealing *in situ* a barcode probe to the amplified nucleic acid molecule. As used herein, the term barcode probe refers to a nucleic acid molecule which can hybridize specifically to a nucleic acid of interest, and which comprises a barcode sequence. As used herein, the terms barcode, barcode sequence and like terms refer to a unique identifier sequence on a probe molecule. The unique identifier sequence can be, for example, a nucleic acid sequence which associates the barcode probe with the nucleic acid of interest to which it hybridizes.

The barcode probes presented herein can be any probe which specifically hybridizes to a nucleic acid species of interest and which comprises at least a first and a second decoder sequence. In some embodiments presented herein, the barcode probes are used as a template for further amplification, thus generating a clonal object that comprises many repeating copies of the barcode, as described in greater detail in the disclosure of WO 2011/123246, the entire content of which is incorporated herein by reference in its entirety. As used herein, the terms “clonal object” refers to a particle having a nucleic acid sequence in one or more copies. An exemplary clonal object is a nucleic acid that has been amplified from a target nucleic acid molecule and in some aspects has a single tandemly repeated sequence of the target nucleic acid molecule. Such tandemly repeated sequences may also be separated with non-target nucleic acids, such as primer binding sites, endonuclease recognition sites, nucleotides linked to

affinity ligands or the like. In particular embodiments a clonal object can be a DNA ball, for example, formed by rolling circle amplification. Methods of generating a clonal object are well known to one of skill in the art and exemplary methods are also described herein. As used herein, a “clonal object” can be synthesized using an
5 amplification technique and thus is a type of amplicon. Accordingly, an amplicon is the nucleic acid product of an amplification reaction.

A method for generating an array of amplified nucleic acid sequences can include the step of attaching at least one second universal primer having a second common priming site to a plurality of sample nucleic acid molecules, thereby attaching
10 a first universal primer and a second universal primer to a sample nucleic acid molecule of the plurality of sample nucleic acid molecules. In a particular embodiment, the first universal primer and the second universal primer can be attached to respective ends of each nucleic acid in the plurality of sample nucleic acid molecules by ligation.

In embodiments that include ligation of a first double stranded nucleic acid end
15 to a second double stranded nucleic acid end, the ends to be ligated can be blunt or can have complementary single stranded overhangs. The use of complementary overhangs generally provides an added measure of specificity over blunt end methods because conditions can be used in which non-complementary sequences will not ligate. Further specificity can be attained by partially filling in one overhang end to make it
20 complementary to another end. This fill in method can be used to disfavor unwanted ligation between nucleic acids in a sample that were generated with the same restriction enzyme.

An amplicon typically contains multiple, tandem copies of the circularized nucleic acid molecule of the corresponding sample nucleic acid. That is, each amplicon
25 contains multiple, tandem copies of a single sample nucleic acid molecule, which was circularized. The number of copies can be varied by appropriate modification of the amplification reaction including, for example, varying the number of amplification cycles run, using polymerases of varying processivity in the amplification reaction and/or varying the length of time that the amplification reaction is run, as well as
30 modification of other conditions known in the art to influence amplification yield.

Generally, the number of copies of a nucleic acid in an amplicon is at least 100, 200, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000 and 10,000 copies, and can be varied depending on the particular application. As disclosed herein, one particular form of an amplicon is as a nucleic acid “ball” having desired dimensions. The number
5 of copies of the nucleic acid molecule can therefore provide a desired size of a nucleic acid “ball” or a sufficient number of copies for efficient subsequent analysis of the amplicon, for example, sequencing.

The terms “target nucleic acid,” “target nucleic acid molecule,” “target nucleic acid species” and any grammatical equivalent thereof, refer to nucleic acid molecules
10 that are desired to be detected, sequenced or otherwise analyzed. Any of a variety of desired target nucleic acid molecules can be utilized, including but not limited to exons, or nucleic acid molecules complementary thereto; cDNA molecules, or nucleic acid molecules complementary thereto; untranslated regions (UTRs) or nucleic acids complementary thereto; promoter and/or enhancer regions, or nucleic acid molecules
15 complementary thereto; evolutionary conserved regions (ECRs), or nucleic acid molecules complementary thereto; transcribed genomic regions, or nucleic acid molecules complementary thereto. Any of a variety of methods can be used to obtain targeted nucleic acid molecules, as disclosed herein. Such methods include, but are not limited to, obtaining a targeted nucleic acid molecule using hybridization-extension
20 capture enrichment; using targeted restriction sites, for example, using an oligonucleotide engineered with a hairpin having a Type IIS restriction enzyme site such as a FokI restriction enzyme site and a locus-specific region; using locus-specific hyperbranched rolling circle amplification; using random-locus-specific primer amplification; using multiplex emulsion PCR; using multiplex bridge PCR; using
25 padlock probe amplification; and using mini-libraries from targeted libraries, as disclosed herein.

As used herein, the terms “target nucleic acid sequence,” “sample nucleic acid sequence” and like terms refer to nucleic acid sequences obtained from samples that are desired to be analyzed.

A nucleic acid sample that is amplified, sequenced or otherwise manipulated in a method disclosed herein can be, for example, DNA or RNA. Exemplary DNA species include, but are not limited to, genomic DNA (gDNA), mitochondrial DNA, chloroplast DNA, episomal DNA, viral DNA and copy DNA (cDNA). One non-limiting example
5 of a subset of genomic DNA is one particular chromosome or one region of a particular chromosome. Exemplary RNA species include, without limitation, coding RNA such as messenger RNA (mRNA), and non-coding RNA (ncRNA) such as transfer RNA (tRNA), microRNA (miRNA), small nuclear RNA (snRNA) and ribosomal RNA (rRNA). Further species of DNA or RNA include fragments or portions of the species
10 listed above or amplified products derived from these species, fragments thereof or portions thereof. The methods described herein are applicable to the above species encompassing all or part of the complement present in a cell. For example, using methods described herein the sequence of a substantially complete genome can be determined or the sequence of a substantially complete targeted nucleic acid sequences
15 such as mRNA or cDNA complement of a cell can be determined.

Useful methods for clonal amplification from single molecules include, for example, rolling circle amplification (RCA) (Lizardi et al., Nat. Genet. 19:225-232 (1998), which is incorporated herein by reference) and oligonucleotide ligation assay (OLA) (See generally U.S. Pat. Nos. 7,582,420, 5,185,243, 5,679,524 and 5,573,907;
20 EP 0 320 308 B1; EP 0 336 731 B1; EP 0 439 182 B1; WO 90/01069; WO 89/12696; and WO 89/09835, all of which are incorporated by reference) technologies. The enhanced signal-to-noise ratio provided by clonal amplification provides a great advantage over existing detection methodologies.

In a particularly useful embodiment, amplicons are generated by rolling circle
25 amplification (RCA), which can be used to generate amplicons having multiple copies of a nucleic acid sequence and which can be used to create nucleic acid “balls,” as disclosed herein. It will be understood that these “balls” need not be perfectly spherical and can include other globular or packed conformations. In a particular embodiment, RCA is primed using the at least one universal primer attached to the sample nucleic
30 acid molecule.

Barcode probes used in the methods presented herein can include padlock probes. Padlock probes are specialized ligation probes and are known in the art, as exemplified generally by Nilsson et al, 1994, Science 265:2085, hereby incorporated by reference. These probes have a first ligation domain that is identical to a first ligation probe, in that it hybridizes to a first target sequence domain, and a second ligation domain, identical to the second ligation probe, that hybridizes to an adjacent target sequence domain. Again, as for OLA, the detection position can be either at the 3' end of the first ligation domain or at the 5' end of the second ligation domain. However, the two ligation domains are connected by a linker, frequently nucleic acid. The configuration of the system is such that upon ligation of the first and second ligation domains of the padlock probe, the probe forms a circular probe, and forms a complex with the target sequence wherein the target sequence is "inserted" into the loop of the circle.

Specific padlock probes are useful for the methods provided herein and include, for example, padlock probes for genomic analyses, as exemplified by Gore, A. et al. Nature 471, 63-67 (2011); Porreca, G.J. et al. Nat Methods 4, 931-936 (2007); and Li, J.B. et al. Genome Res 19, 1606-1615 (2009), the content of each of which is incorporated by reference in its entirety.

Other padlock probes useful for the methods provided herein include, for example, padlock probes for transcriptomic analyses, as exemplified by Zhang, K. et al. Nat Methods 6, 613-618 (2009); Noggle, S. et al. Nature 478, 70-75 (2011); and Li, J.B. et al. Science 324, 1210-1213 (2009), the content of each of which is incorporated by reference in its entirety.

Other examples of padlock probes useful for the methods provided herein include, for example, padlock probes for methylomic analyses, as exemplified by Deng, J. et al. Nat Biotechnol 27, 353-360 (2009); Hansen, K.D. et al. Nat Genet 43, 768-775 (2011); Diep, D. et al. Nature Methods (2012), the content of each of which is incorporated by reference in its entirety.

As disclosed herein, the amplicons can be compacted prior to hybridizing or binding to a bead described herein. Methods of compacting amplicons are known in the

art (for example, as described by Bloomfield, Curr. Opin. Struct. Biol. 6(3): 334-41 (1996), and Drmanac et al., US 2007/0099208 A1, each of which is incorporated herein by reference) and disclosed herein. For example, an alcohol or polyamine such as spermine or spermidine can be used. A compacted nucleic acid will have a structure that is more densely packed than the structure of the nucleic acid in the absence of a compacting agent or compacting condition and the structure will typically resemble a ball or globule. The generation of such compacted nucleic acid balls is useful for fabricating one clonal object bound to one bead, as discussed herein in more detail. Various methods can be used to generate balls of a desired size, for example, using various compacting techniques and/or varying the number of copies in an amplicon. Generally, the compacted amplicons have an average diameter or width ranging from about 0.1 μm to about 5 μm , for example, about 0.1 μm , about 0.2 μm , about 0.5 μm , about 1 μm , 2 μm , about 3 μm , about 4 μm and about 5 μm .

15 Decoder Sequences and Decoder Probes

The methods and compositions presented herein make use of unique barcode probes having decoder sequences to identify multiple nucleic acid targets in a multiplex fashion. The instant innovations overcome limitations of existing colony identification methodologies by surprisingly and successfully applying decoding strategies that entail sequential hybridizations of labeled decoder probes to decoder sequences found in immobilized nucleic acids within an individual cell. Such decoding strategies using decoder sequences have until now been used for decoding of randomly assembled arrays of beads. Thus, the decoding strategies presented herein are flexible and scalable and can be used to detect and decode thousands of unique amplified nucleic acid species.

In some embodiments, a barcode probe can comprise at least a first decoder sequence and a second decoder sequence. In some embodiments, a barcode probe can comprise 2, 3, 4, 5, 6, 7, 8, 9, 10 or more than 10 decoder sequences. As used herein, the terms decoder sequence, stage, slot and like terms refer to a unique sequence which is part of a plurality of unique sequences are designed to each specifically hybridize to a

labeled decoder probe. During decoding, labeled decoder probe oligonucleotides are sequentially hybridized to the decoder sequences to specifically determine the identity of each barcode sequence. Sequential hybridization of decoder probes to decoder sequences is described by Gunderson et al. (Genome Res. (2004) 14:870-877), the content of which is incorporated by reference in its entirety. Additional description of decoding algorithms and the use of sequential hybridization useful in the methods presented herein can be found in U.S. Patent 7,060,431; U.S. Patent 7,033,754; U.S. Patent 7,226,734; and U.S. Patent 7,455,971, the disclosures of each of which are hereby incorporated by reference in their entireties. The principles of sequential hybridization to decode bead arrays are surprisingly robust and effective when applied to amplified nucleic acids *in situ* according to the methods provided herein.

As an example, one or more target nucleic acid species can be amplified, either specifically or non-specifically, to generate an amplified molecule containing many repeats of the amplified nucleic acid in tandem. In the embodiment shown in Fig. 1, cDNA is generated *in situ*, using either random hexamers or oligo dT primers or both. Then the cDNA molecules are amplified using rolling circle amplification (RCA) to generate a “DNA ball” or “rolony”. The amplified nucleic acids can then be hybridized with one or more unique padlock probes with unique molecular barcodes. When a padlock probe specifically hybridizes to a sequence of interest, the probe is ligated to form, for example, a circularized barcoded padlock probe. The amplified nucleic acid molecules and/or the hybridized padlock probes are immobilized within the cell using the crosslinking methods described herein, and decoding of the barcode probe takes place by multiple rounds of sequential hybridization. In the example shown in Fig. 1, decoding takes place via a set of barcode probes having decoder sequences that can be decoded by eight rounds of sequential hybridization with fluorescently labeled oligos and imaging of fluorescent images.

The decoding strategies presented herein provide several key advantages over existing sequence identification methods. For example, one advantage is the simplicity and robustness of the decoding process, which involves hybridization and washing of oligo pools at room temperature. As such, heating and cooling steps are not necessary,

thus minimizing damage to the cellular structure and reducing the potential drift of the specimen throughout the imaging process. As an additional advantage, the signals that are generated are more even and have higher intensity inside a structured tissue than methods that rely on other detection techniques. Further, in many embodiments, only
 5 two fluors are needed for decoding, which provides convenient spectral distinction using simple detection devices and leaves enough spectrum for labeling cellular structure using other fluors. Of course more than two colors can be used to expand the number of sequences that are decoded, using modifications of the method set forth in further detail elsewhere herein and in the references that are incorporated by reference
 10 above as describing sequential hybridization of decoder probes to decoder sequences.

The at least first and second decoding sequences can be any suitable length to ensure specific hybridization. Thus, for example, in some embodiments, each decoding sequence can be 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or greater than 20 nucleotides in length.

15 In some embodiments, the methods provided herein allow for detection of a single target nucleotide species of interest. In other embodiments, the methods allow for multiplex detection of a plurality of nucleotide species of interest. The use of barcode probes with unique identifier sequences as described above allows for simultaneous detection of 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100,
 20 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, 10,000 or more than 10,000 unique nucleic acid target molecules within a single cell. Thus, in contrast to existing *in situ* detection methods, which are limited by the requirement for unique labels (such as fluorescent labels) for each probe type, the methods presented herein
 25 have the advantage of virtually limitless numbers of individually detected molecules in parallel and *in situ*.

Immobilization of nucleic acids *in situ*

In embodiments presented herein, amplified nucleic acid molecules, barcode
 30 probes and other nucleic acid molecules are immobilized within the structure of the cell

by crosslinking the nucleic acid to structural proteins in the cell. Any suitable method for immobilizing nucleic acid molecules within a cell may be used in the methods and compositions presented herein.

5 Detection of Disease-Causing Mutations and/or Common Variants with Specific Padlock Probes

The methods and compositions provided herein make it possible to specifically detect one or more disease-causing mutations, alleles or common variants *in situ* within an individual cell or plurality of cells. Thus in some embodiments, the methods make
10 use of one or more specific barcode probes which are correlated with a disease-causing mutation or mutations, alleles or common variants. In some aspects, the method can comprise ligation of a mutation-specific barcode probe, such as a padlock probe, only when the disease-causing mutation, allele or common variant is present in the nucleic acid molecule.

Any mutation, allele or common variant type can be a target for a mutation-specific barcode probe. The disease-causing mutation, allele or common variant can be, for example, a base substitution, an insertion mutation, a deletion mutation, a gene amplification, a gene deletion, a gene fusion event or gene inversion event or any other mutation which is associated with one or more phenotypes of interest. In some
20 embodiments, the mutation, allele or common variant is associated with an increased predisposition for one or more diseases or other phenotypes. In some embodiments, the mutation, allele or common variant is associated with a decreased predisposition for one or more diseases or other phenotypes. For example, some mutations, alleles or common variants are associated with a cancer phenotype, such as decreased growth inhibition,
25 evasion of immune detection, de-differentiation, and the like. Somatic mutations that can be detected using the method provided herein include for example, mutations to BRAF, EGFR, Her2/ERBB2, and other somatic mutations as exemplified by Greenman et al., Nature (2007) 446:153-158, hereby incorporated by reference in its entirety. Examples of specific barcodes for detection of somatic mutations are provided in the
30 Example section below.

Detection of mutation-specific barcode probes can further comprise evaluating a signal to determine for example, the presence, quantity, frequency, and/or spatial distribution of a disease-causing mutation, allele or common variant within a tissue.

Probes as described herein can also be designed to target specific genomic regions, such as, for example, the specific microsatellite repeat on the X chromosome. In addition, probes can be used as internal controls, either as a positive or negative control, or as a standard for normalization of one or more other signals during detection.

Probes as described herein are also useful for detection of methylation patterns *in situ*. In such embodiments, nucleic acids within the cell or cells of interest may be subjected to bisulfite conversion as is known in the art. Following bisulfite conversion, one or more methylation-specific barcode probes may be specifically hybridize to a methylation pattern of interest and then amplified for decoding and detection as described above.

15 Cell Handling for *In Situ* Detection and Decoding

The methods and compositions make use of hybridization, amplification, washing and detection within individual cells. These methods are performed *in situ* on isolated cells or in tissue sections that have been prepared according to methodologies known in the art. Methods for permeabilization and fixation of cells and tissue samples are well-known in the art, as exemplified by Cremer et al., *The Nucleus: Volume 1: Nuclei and Subnuclear Components*, R. Hancock (ed.) 2008; and Larsson et al., *Nat. Methods* (2010) 7:395-397 and associated supplemental materials, the content of each of which is incorporated herein by reference in its entirety. Any suitable permeabilization and/or fixation technologies can be used for making cells and the target nucleic acid molecules therein available for the amplification and detection methods provided herein.

Detection of Signals

The signals generated by fluorescent imaging of one or more probes during decoding, sequencing and other detection methods carried out *in situ* and can be

detected using any suitable imaging and detection technologies which allow for detection of individual amplicons *in situ*. Suitable imaging technologies are known in the art, as exemplified by Larsson et al., Nat. Methods (2010) 7:395-397 and associated supplemental materials, the entire content of which is incorporated by reference herein
5 in its entirety.

An example demonstrating *in situ* imaging and 3D reconstruction of *in situ* detection is set forth in Figure 1 and Example 1.

Sequencing to Identify Barcode Probe

10 In some embodiments, the barcode sequence in the barcode probe may be detected and/or decoded *in situ* by sequencing methodologies, such as direct sequencing, including sequencing by synthesis, sequencing by hybridization, and the like.

One example is sequencing-by-synthesis (SBS). In SBS, extension of a nucleic
15 acid primer along a nucleic acid template (e.g. a target nucleic acid or amplicon thereof) is monitored to determine the sequence of nucleotides in the template. The underlying chemical process can be polymerization (e.g. as catalyzed by a polymerase enzyme). In a particular polymerase-based SBS embodiment, fluorescently labeled nucleotides are added to a primer (thereby extending the primer) in a template dependent fashion such
20 that detection of the order and type of nucleotides added to the primer can be used to determine the sequence of the template. A plurality of different templates at different sites within an individual cell as set forth herein can be subjected to an SBS technique under conditions where events occurring for different templates can be distinguished due to their location within the individual cell.

25 Flow cells provide a convenient format for housing an individual cell (or a spatially arrayed collection of individually resolvable cells) having amplified nucleic acid molecules produced by the methods of the present disclosure. One or more individual cells in such a format can be subjected to an SBS or other detection technique that involves repeated delivery of reagents in cycles. For example, to initiate
30 a first SBS cycle, one or more labeled nucleotides, DNA polymerase, etc., can be

flowed into/through a flow cell that houses one or more individual cells having amplified nucleic acid templates. Those sites of an individual cell where primer extension causes a labeled nucleotide to be incorporated can be detected. Optionally, the nucleotides can further include a reversible termination property that terminates further primer extension once a nucleotide has been added to a primer. For example, a nucleotide analog having a reversible terminator moiety can be added to a primer such that subsequent extension cannot occur until a deblocking agent is delivered to remove the moiety. Thus, for embodiments that use reversible termination, a deblocking reagent can be delivered to the flow cell (before or after detection occurs). Washes can be carried out between the various delivery steps. The cycle can then be repeated n times to extend the primer by n nucleotides, thereby detecting a sequence of length n . Exemplary SBS procedures, fluidic systems and detection platforms that can be readily adapted for use with amplicons produced *in situ* by the methods of the present disclosure are described, for example, in Bentley et al., *Nature* 456:53-59 (2008), WO 04/018497; US 7,057,026; WO 91/06678; WO 07/123744; US 7,329,492; US 7,211,414; US 7,315,019; US 7,405,281, and US 2008/0108082, each of which is incorporated herein by reference.

Other sequencing procedures that use cyclic reactions can be used, such as pyrosequencing. Pyrosequencing detects the release of inorganic pyrophosphate (PPi) as particular nucleotides are incorporated into a nascent nucleic acid strand (Ronaghi, et al., *Analytical Biochemistry* 242(1), 84-9 (1996); Ronaghi, *Genome Res.* 11(1), 3-11 (2001); Ronaghi et al. *Science* 281(5375), 363 (1998); US 6,210,891; US 6,258,568 and US 6,274,320, each of which is incorporated herein by reference). In pyrosequencing, released PPi can be detected by being immediately converted to adenosine triphosphate (ATP) by ATP sulfurylase, and the level of ATP generated can be detected via luciferase-produced photons. Thus, the sequencing reaction can be monitored via a luminescence detection system. Excitation radiation sources used for fluorescence based detection systems are not necessary for pyrosequencing procedures. Useful fluidic systems, detectors and procedures that can be adapted for application of pyrosequencing to amplicons produced *in situ* according to the present disclosure are

described, for example, in WIPO Pat. App. Ser. No. PCT/US11/57111, US 2005/0191698 A1, US 7,595,883, and US 7,244,559, each of which is incorporated herein by reference.

Some embodiments can utilize methods involving the real-time monitoring of DNA polymerase activity. For example, nucleotide incorporations can be detected through fluorescence resonance energy transfer (FRET) interactions between a fluorophore-bearing polymerase and γ -phosphate-labeled nucleotides, or with zeromode waveguides (ZMWs). Techniques and reagents for FRET-based sequencing are described, for example, in Levene et al. *Science* 299, 682–686 (2003); Lundquist et al. *Opt. Lett.* 33, 1026–1028 (2008); Korlach et al. *Proc. Natl. Acad. Sci. USA* 105, 1176–1181 (2008), the disclosures of which are incorporated herein by reference.

Some SBS embodiments include detection of a proton released upon incorporation of a nucleotide into an extension product. For example, sequencing based on detection of released protons can use an electrical detector and associated techniques that are commercially available from Ion Torrent (Guilford, CT, a Life Technologies subsidiary) or sequencing methods and systems described in US 2009/0026082 A1; US 2009/0127589 A1; US 2010/0137143 A1; or US 2010/0282617 A1, each of which is incorporated herein by reference. Methods set forth herein for amplifying target nucleic acids using kinetic exclusion can be readily applied to substrates used for detecting protons. More specifically, methods set forth herein can be used to produce clonal populations of amplicons *in situ* that are used to detect protons.

Exemplary methods for array-based expression and genotyping analysis that can be applied to *in situ* detection according to the present disclosure are described in US Pat. Nos. 7,582,420; 6,890,741; 6,913,884 or 6,355,431 or US Pat. Pub. Nos. 2005/0053980 A1; 2009/0186349 A1 or US 2005/0181440 A1, each of which is incorporated herein by reference.

An advantage of the methods set forth herein is that they provide for rapid and efficient detection of a plurality of target nucleic acid in parallel. Accordingly the present disclosure provides integrated systems capable of preparing and detecting nucleic acids *in situ* using techniques known in the art such as those exemplified above.

Thus, an integrated system of the present disclosure can include fluidic components capable of delivering amplification reagents to one or more individual cells, the system comprising components such as pumps, valves, reservoirs, fluidic lines and the like. A particularly useful fluidic component is a flow cell. A flow cell can be configured and/or used in an integrated system for *in situ* detection of target nucleic acids. Exemplary flow cells are described, for example, in US 2010/0111768 A1 and US Ser. No. 13/273,666, each of which is incorporated herein by reference. As exemplified for flow cells, one or more of the fluidic components of an integrated system can be used for an amplification method and for a detection method. Taking a nucleic acid sequencing embodiment as an example, one or more of the fluidic components of an integrated system can be used for an amplification method set forth herein and for the delivery of sequencing reagents in a sequencing method such as those exemplified above. Alternatively, an integrated system can include separate fluidic systems to carry out amplification methods and to carry out detection methods. Examples of integrated sequencing systems that are capable of creating amplified nucleic acids and also determining the sequence of the nucleic acids include, without limitation, the MiSeq platform (Illumina, Inc., San Diego, CA) and devices described in US Ser. No. 13/273,666, which is incorporated herein by reference.

A system capable of carrying out a method set forth herein need not be integrated with a detection device. Rather, a stand-alone system or a system integrated with other devices is also possible. Fluidic components similar to those exemplified above in the context of an integrated system can be used in such embodiments.

A system capable of carrying out a method set forth herein, whether integrated with detection capabilities or not, can include a system controller that is capable of executing a set of instructions to perform one or more steps of a method, technique or process set forth herein. For example, the instructions can direct the performance of steps for creating a set of amplicons *in situ*. Optionally, the instructions can further direct the performance of steps for detecting nucleic acids using methods set forth previously herein. A useful system controller may include any processor-based or microprocessor-based system, including systems using microcontrollers, reduced

instruction set computers (RISC), application specific integrated circuits (ASICs), field programmable gate array (FPGAs), logic circuits, and any other circuit or processor capable of executing functions described herein. A set of instructions for a system controller may be in the form of a software program. As used herein, the terms
5 “software” and “firmware” are interchangeable, and include any computer program stored in memory for execution by a computer, including RAM memory, ROM memory, EPROM memory, EEPROM memory, and non-volatile RAM (NVRAM) memory. The software may be in various forms such as system software or application software. Further, the software may be in the form of a collection of separate programs,
10 or a program module within a larger program or a portion of a program module. The software also may include modular programming in the form of object-oriented programming.

EXAMPLE 1

15 Design and optimization of a set of padlock probes and hybridization oligos for targeted detection and quantification of a subset of genes.

This example presents an innovative method based on padlock circularization and hybridization-based decoding (Fig. 1). This method is built upon ultra-high degree multiplexed capture (>300,000 per reaction) using padlock probes for useful for
20 genomic, transcriptomic and methylomic analyses, as well as technology for accurate decoding of randomly assembled bead arrays.

In some embodiments, colonies for an entire transcriptome can be generated using either random hexamers or oligo-dT primers (Fig. 1A). In some embodiments, colonies for a specific subset of genes of interest can then be tagged with gene-specific
25 barcodes using a library of padlock probes (Fig. 1B).

Padlock probes are designed in a similar manner as the full transcriptome probe sets previously developed for digital RNA allelotyping according to the incorporated materials disclosed in Zhang, K. et al. Nat Methods 6, 613-618 (2009) and Noggle, S. et al. Nature 478, 70-75 (2011). A neural network, trained for predicting probe efficiency
30 using experimental data accumulated on over 400,000 probes, is integrated into a probe

design algorithm called ppDesigner. The entire probe design, synthesis and production process has been highly optimized.

Long oligonucleotides (up to 300nt in length) acquired from Agilent Technologies contain amplification adaptors to allow production of essentially an unlimited amount of functional padlock probes from a set of seed oligos. The design of the barcoding region incorporates the error-checking scheme described in the incorporated materials of Gunderson et al. (Genome Res. (2004) 14:870-877). For up to ~1700 target genes, eight rounds of hybridization decoding are performed. In each round, a colony can have three states: two fluorescent (FAM and Cy3) “ON” states and one nonfluorescent OFF state. In this three-state, eight-stage decoding scheme, the valid barcode for each gene is required to have exactly two OFF stages and six ON stages for error checking. Each padlock probe only needs six slots for hybridization (corresponding to the six ON stages). For the hybridization sequences used in these six slots, a set of 16 (8 stages x 2 ON states) 20-mers is selected that are orthogonal to the human genome and have uniform annealing parameters. This eight-stage barcoding scheme is sufficient to decode up to 1792 transcripts with a low error rate ($<10^{-4}$).

EXAMPLE 2

Design and optimization of a set of padlock probes for targeted detection and quantification of BRAF V600E, EGFR T790M and EGFR L858R mutations

In this example, allele-specific (padlock) probes are generated for disease causing mutations. Specifically, point mutations are detected in genomic DNA and the decoding scheme is utilized to measure the frequency and spatial distribution of these mutations within tissue sections at single cell resolution. For each of the specifically targeted mutations described below, a padlock probe is designed for specific hybridization to either the wild-type or mutant version of the gene. Then, the padlock probe serves as a template for RCA as described in greater detail herein. The resulting amplicon is detected and decoded using a barcode sequence integrated into the padlock probe.

In an alternative embodiment, the first padlock probe does not comprise a barcode sequence, and a second set of padlock probes is hybridized to the amplicons, and decoder sequences found in the second set of padlock probes are decoded through sequential hybridization as described above.

Probe for detection of BRAF V600E mutation

5' -AAGTATGGGACCCACTCCATCG [A/T] CTGTAGTAGACCAAAATCACCTA-3'
5' -CCGAGACACTGTTCAATCATATGGGACCCACTCCATCCAGATTTC [A/T] CTGTAGTAGACCAAAATCACCTATTTTACTGTGAGGCTTCATG-3'

Wild-type probe (10101):

CGATGGAGTGGTCCCATCAGTTATCGCGCTCGATTGGCAGTAGAACTTGC GCGGATACACGGTCGAGGTCGAGCTCGGAAATACGTCGTGCTACTGCGTAGGTGATTTGGTCTAGCTACAG⁷

Mutant Probe (2022):

CGATGGAGTGGTCCCATCAGTTACGTATCGGTAGTCGCAACGCATCTACTTCGTCCGCTCAGACCATCGAGGTCGAGCTACCATTCGCGATTCGATTCATAGGTGATTTGGTCTAGCTACAG⁸

15

Probe for detection of EGFR T790M mutation

20 CCACCGTGCAGTCATCA [C/T] CATGCCCTTGGCTGCCTCCT
TGGCGCCTGTGGGCATCTGCCCTACCTCCACCGTGCAGTCATCA [C/T] CGAGCTCATGCCCTTGGCGTGCCTCGGACATGTCCGGGAACA

Wild-type probe: 01011

CATGCCCTTGGCTGGCTCCTCTTGGTGGGATACGGAGTACGAATTGTCGCGCTCTACGATTCGAGGTCGACTCGGAATACGTGTTGACTGGCCACCGTGCAGCTCATCAC

25 Mutant Probe: 02022³³

CATGCCCTTCGGCTGCCTCCTCAACGGTATTCCGGTCGTATCCCCGTTTGATCGTTCGAGGTCGAGATCGAGCTCCGACCTACCAATTCGGGTGCGATTCCACACCGTGCAGCTCATTCAT

30

Probe for detection of EGFR L858R mutation

GTCAAGATCACAGATTTTGGGC [T/G] CTGCTGGGTGCGGAAGAGAAAAGAATA
 GTACTGGTGTAAACACCGCAGCATGTCAAGATCCACAGATTTTGGGC [T/G] GGCCAAACTGCTGGGTGCGGAAGAGCAAGAATACCATGCCAGAAGG

5 Wild-type probe: 00112

CTGCTGGGTGCGGAAGAGAAAAGAATATAGAACTTGCGCGGATACACGGCGAATTGTCCGCGCTCTACGATTTCGAGGTCCGACCTACCATTCGCGTGCGATTCCAGT
 CAAGATCACAGATTTTGGGC^{***}

Mutant Probe: 02201

CTGCTGGGTGCGGAAGAGAAAAGAATACAACGGTATTTCGGTTCGTCA^{***}TCCCTCTACTTCGTCCGCTCAGACCATCGAGGTCCGACTCGGAATACGTCGTTGACTGCGGT
 CAAGATCACAGATTTTGGGC^{***}

10

EXAMPLE 3

Design and optimization of a set of padlock probes for targeted detection and quantification of Her2/ERBB2 amplification and/or deletion

In this example, allele-specific (padlock) probes are generated for
5 Her2/ERBB2 mutations. Specifically, gene amplification and/or deletion is detected in genomic DNA and the decoding scheme is utilized to measure the frequency and spatial distribution of gene amplifications and/or deletions within tissue sections at single cell resolution.

Design, hybridization and detection are performed as described above in
10 Example 2. A specifically designed padlock probe is described below.

Probe for detection of Her2/ERBB2

AGCTCTTTGAGGACAACTATGCGCTG AGACAATGGAGACCCGCTGAAC
TGTGGAGGCACCCAGCTCTTTTCAGGACAACCTATGCCCCGCTGGCCGCTAGACAATGGAGACCCGCTGGACAATACCCCTGTCTCACAGGGGCTCCCCA
Wild-type probe: 20022
AGACAATGGAGACCCGCTGAACACGTATCGGTAGTCGCAACGCACCCGTTTGATCGTTTCGACCCGAGATCGAGGTCCGACCTACCATTCGCGTGCATTCCAAGCTCT
TTGAGGACAACTATGCGCTG

EXAMPLE 4

Design and optimization of a set of padlock probes for targeted detection and quantification of microsatellite repeat

In this example, allele-specific (padlock) probes are generated for detection of
5 chrX microsatellite repeats. Specifically, the specific microsatellite repeats are detected
in genomic DNA and the decoding scheme is utilized to evaluate microsatellite repeats
within tissue sections at single cell resolution.

Design, hybridization and detection are performed as described above in
Example 2. A specifically designed padlock probe is described below.

5 Wild-type probe: 11010

ATGTATCGGCTCGATTGGCAGTAGAACTTGC GCGGATACACGG
GTGCTGCCAGGACTTCTTGCATGTATCGGCTCGATTGGCAGGTCGATACGAGTATCGAGGTCCGACCGAATTGTCCGCGCTCTACGATCCTGCTCTCTCTGTCTCAAC

Throughout this application various publications, patents and/or patent applications have been referenced. The disclosure of these publications in their entireties is hereby incorporated by reference in this application.

The term comprising is intended herein to be open-ended, including not only the
5 recited elements, but further encompassing any additional elements.

A number of embodiments have been described. Nevertheless, it will be understood that various modifications may be made. Accordingly, other embodiments are within the scope of the following claims.

10

What is claimed is:

1. A method of obtaining sequence information from an individual cell comprising:

- 5 (a) amplifying a nucleic acid molecule *in situ* within a cell, thereby generating an amplified nucleic acid molecule;
- (b) annealing *in situ* a barcode probe comprising at least a first decoder sequence and a second decoder sequence to said amplified nucleic acid molecule;
- (c) hybridizing a first labeled decoder probe to said first decoder
10 sequence;
- (d) detecting a signal indicative of said hybridizing;
- (e) hybridizing a second labeled decoder probe to said second decoder sequence; and
- (f) detecting a signal indicative of said hybridizing, thereby
15 generating sequence information regarding said amplified nucleic acid molecule.

2. A method of obtaining sequence information regarding a plurality of nucleic acid molecules from an individual cell comprising:

- (a) amplifying a plurality of nucleic acid molecules *in situ* within a cell, thereby generating a set of amplified nucleic acid molecules;
- 20 (b) for each of a plurality of amplified nucleic acid molecules in said set, annealing *in situ* a barcode probe comprising at least a first decoder sequence and a second decoder sequence to said amplified nucleic acid molecule; and
- (c) hybridizing a first labeled decoder probe to said first decoder
25 sequence;
- (d) detecting a plurality of signals indicative of said hybridizing;
- (e) hybridizing a second labeled decoder probe to said second decoder sequence; and

(f) detecting a plurality of signals indicative of said hybridizing, thereby generating sequence information regarding said plurality of amplified nucleic acid molecules.

3. The method of any of claims 1-2, wherein said first labeled decoder
5 probe is removed prior to hybridizing said second labeled decoder probe.

4. The method of any of claims 1-2, wherein detecting comprises sequencing at least said first decoder sequence *in situ*.

5. The method of claim 4, further comprising sequencing at least said second decoder sequence *in situ*.

10 6. The method of any of claims 1-2, wherein said amplifying comprises rolling circle amplification.

7. The method of any of claims 1-2, wherein said barcode probe comprises a rolling circle amplification probe.

8. The method of claim 7, wherein said rolling circle amplification probe
15 comprises a padlock probe.

9. The method of any of claims 1-2, wherein said barcode probe comprises an oligonucleotide ligation assay probe.

10. The method of any of claims 1-2, wherein said amplified nucleic acid molecule or molecules are immobilized within said cell.

20 11. The method of any of claims 1-2, wherein said amplified nucleic acid molecule is amplified from an RNA template.

12. The method of claim 11, further comprising cDNA synthesis using said RNA as a template.

13. The method of any of claims 1-2, wherein said amplified nucleic acid
25 molecule comprises DNA.

14. The method of any of claims 1-2, wherein said nucleic acid molecule comprises a disease-causing mutation.

15. The method of claim 13, comprising hybridizing a barcode probe which is correlated with said disease-causing mutation or common variant.

16. The method of claim 15, comprising ligation of a mutation-specific padlock probe only when said disease-causing mutation or common variant is present in said nucleic acid molecule.

17. The method of claim 15, wherein said disease-causing mutation
5 comprises a base substitution.

18. The method of claim 15, wherein said disease-causing mutation comprises an insertion mutation.

19. The method of claim 15, wherein said disease-causing mutation comprises a deletion mutation.

10 20. The method of any of claims 15, wherein said disease-causing mutation comprises a gene amplification.

21. The method of claim 15, wherein said disease-causing mutation comprises a gene deletion.

15 22. The method of claim 15, wherein said disease-causing mutation comprises a gene fusion event or gene inversion event.

23. The method of claim 15, wherein detecting further comprises evaluating a signal to determine quantity of said disease-causing mutation within a tissue.

24. The method of claim 15, wherein detecting further comprises evaluating a signal to determine frequency of said disease-causing mutation within a tissue.

20 25. The method of claim 15, wherein detecting further comprises evaluating a signal to determine spatial distribution within a tissue of a gene amplification related to said disease-causing mutation.

26. The method of claim 15, wherein detecting further comprises evaluating a signal to determine the presence of a disease-causing mutation within a tissue.

25 27. A method of obtaining sequence information from an individual cell comprising:

(a) annealing a barcode probe specific for a disease-causing mutation to a nucleic acid molecule *in situ* within a cell, said barcode probe comprising at least a first decoder sequence and a second decoder sequence;

(b) amplifying said barcode probe *in situ*, thereby generating an amplified barcode probe;

(c) hybridizing a first labeled decoder probe to said first decoder sequence;

5 (d) detecting a signal indicative of said hybridizing;

(e) hybridizing a second labeled decoder probe to said second decoder sequence; and

(f) detecting a signal indicative of said hybridizing, thereby generating sequence information regarding said amplified nucleic acid molecule.

10 28. The method of claim 27, wherein said first labeled decoder probe is removed prior to hybridizing said second labeled decoder probe.

29. The method of claim 27, wherein detecting comprises sequencing at least said first decoder sequence *in situ*.

15 30. The method of claim 29, further comprising sequencing at least said second decoder sequence *in situ*.

31. The method of claim 27, wherein said amplifying comprises rolling circle amplification.

32. The method of claim 31, wherein said barcode probe comprises a rolling circle amplification probe.

20 33. The method of claim 32, wherein said rolling circle amplification probe comprises a padlock probe.

34. The method of claim 27, wherein said barcode probe comprises an oligonucleotide ligation assay probe.

25 35. The method of claim 27, wherein said amplified nucleic acid molecule or molecules are immobilized within said cell.

36. The method of claim 27, wherein said amplified nucleic acid molecule is amplified from an RNA template.

37. The method of claim 36, further comprising cDNA synthesis using said RNA as a template.

38. The method of claim 27, wherein said amplified nucleic acid molecule comprises DNA.

39. The method of claim 27, wherein said nucleic acid molecule comprises a disease-causing mutation.

5 40. The method of claim 39, comprising hybridizing a barcode probe which is correlated with said disease-causing mutation or common variant.

41. The method of claim 39, comprising ligation of a mutation-specific padlock probe only when said disease-causing mutation or common variant is present in said nucleic acid molecule.

10 42. The method of claim 39, wherein said disease-causing mutation comprises base substitution.

43. The method of claim 39, wherein said disease-causing mutation comprises an insertion mutation.

15 44. The method of claim 39, wherein said disease-causing mutation comprises a deletion mutation.

45. The method of claim 39, wherein said disease-causing mutation comprises a gene amplification.

46. The method of claim 39, wherein said disease-causing mutation comprises a gene deletion.

20 47. The method of claim 39, wherein said disease-causing mutation comprises a gene fusion event or gene inversion event.

48. The method of claim 39, wherein detecting further comprises evaluating a signal to determine quantity of said disease-causing mutation within a tissue.

25 49. The method of claim 39, wherein detecting further comprises evaluating a signal to determine frequency of said disease-causing mutation within a tissue.

50. The method of claim 39, wherein detecting further comprises evaluating a signal to determine spatial distribution within a tissue of a gene amplification related to said disease-causing mutation.

30 51. The method of claim 39, wherein detecting further comprises evaluating a signal to determine the presence of a disease-causing mutation within a tissue.

52. A method of simultaneous detection of nucleic acids in several individual cells comprising:

- (a) amplifying a plurality of nucleic acid molecules *in situ* within a plurality of cells, thereby generating a set of amplified nucleic acid molecules;
- 5 (b) for each of a plurality of amplified nucleic acid molecules in said set, annealing *in situ* a barcode probe comprising at least a first decoder sequence and a second decoder sequence to said amplified nucleic acid molecule; and
- (c) hybridizing a first labeled decoder probe to said first decoder
10 sequence;
- (d) detecting a plurality of signals indicative of said hybridizing;
- (e) hybridizing a second labeled decoder probe to said second decoder sequence; and
- (f) detecting a plurality of signals indicative of said hybridizing,
15 thereby generating sequence information regarding said plurality of amplified nucleic acid molecules.

53. A method of simultaneous detection of nucleic acids in several individual cells comprising:

- (a) annealing a barcode probe specific for a disease-causing mutation
20 to a nucleic acid molecule *in situ* within a plurality of cells, said barcode probe comprising at least a first decoder sequence and a second decoder sequence;
- (b) amplifying said barcode probe *in situ*, thereby generating an amplified barcode probe;
- (c) hybridizing a first labeled decoder probe to said first decoder
25 sequence;
- (d) detecting a signal indicative of said hybridizing;
- (e) hybridizing a second labeled decoder probe to said second decoder sequence; and
- (f) detecting a signal indicative of said hybridizing, thereby
30 generating sequence information regarding said amplified nucleic acid molecule.

54. The method of any of claims 52-53, wherein the plurality of cells comprises a pool of cell populations from blood or any liquid biospecimen.

55. The method of any of claims 52-53, wherein the plurality of cells comprises cells from a tissue section.

5 56. The method of claim 55, wherein original spatial relationships of the cells are retained.

57. A method of obtaining sequence information from one or more individual cells comprising:

10 (a) annealing *in situ* a barcode probe comprising at least a first decoder sequence and a second decoder sequence to a target nucleic acid molecule;

(b) hybridizing a first labeled decoder probe to said first decoder sequence;

(c) detecting a signal indicative of said hybridizing;

15 (d) hybridizing a second labeled decoder probe to said second decoder sequence; and

(e) detecting a signal indicative of said hybridizing, thereby generating sequence information regarding said target nucleic acid molecule.

ABSTRACT OF THE DISCLOSURE

Presented are methods and compositions for obtaining sequence information from one or more individual cells. The methods are useful for obtaining sequence
5 information for a single nucleotide sequence, and for multiplex generation of sequence information from one or more individual cells.

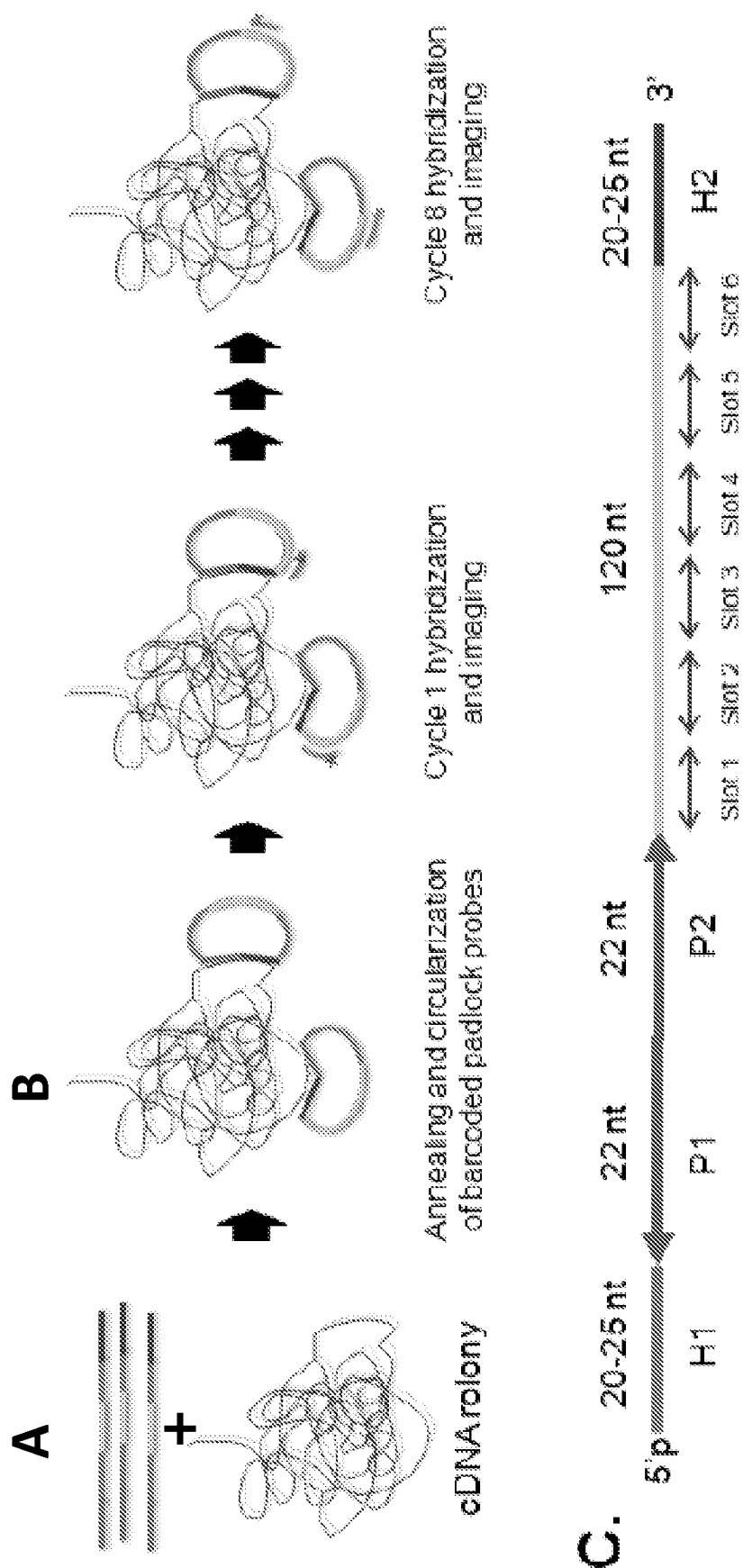


FIG. 1

Electronic Patent Application Fee Transmittal

Application Number:

Filing Date:

Title of Invention:

METHODS AND COMPOSITIONS FOR SINGLE CELL GENOMICS

First Named Inventor/Applicant Name:

Jian-Bing Fang

Filer:

Brent C. Moore/Chelsea Pearsall

Attorney Docket Number:

IP-0645-PRV-A

Filed as Large Entity

Provisional Filing Fees

Description

Fee Code

Quantity

Amount

**Sub-Total in
USD(\$)**

Basic Filing:

Provisional Application Filing

1005

1

260

260

Pages:

Claims:

Miscellaneous-Filing:

Petition:

Patent-Appeals-and-Interference:

Post-Allowance-and-Post-Issuance:

Extension-of-Time:

Description	Fee Code	Quantity	Amount	Sub-Total in USD(\$)
Miscellaneous:				
Total in USD (\$)				260

Electronic Acknowledgement Receipt

EFS ID:	19357952
Application Number:	62014538
International Application Number:	
Confirmation Number:	3783
Title of Invention:	METHODS AND COMPOSITIONS FOR SINGLE CELL GENOMICS
First Named Inventor/Applicant Name:	Jian-Bing Fang
Customer Number:	29389
Filer:	Brent C. Moore/Chelsea Pearsall
Filer Authorized By:	Brent C. Moore
Attorney Docket Number:	IP-0645-PRV-A
Receipt Date:	19-JUN-2014
Filing Date:	
Time Stamp:	17:58:14
Application Type:	Provisional

Payment information:

Submitted with Payment	yes
Payment Type	Deposit Account
Payment was successfully received in RAM	\$ 260
RAM confirmation Number	4664
Deposit Account	502690
Authorized User	

The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:

Charge any Additional Fees required under 37 C.F.R. Section 1.16 (National application filing, search, and examination fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.17 (Patent application and reexamination processing fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.19 (Document supply fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.20 (Post Issuance fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.21 (Miscellaneous fees and charges)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Provisional Cover Sheet (SB16)	IP-0645-PRV-A_ProvisionalSB_signed.pdf	1477315	no	3
			cf1341bfec52bc79ede9cb6ca717fd6aa0604df		
Warnings:					
Information:					
2		IP-0645-PRV_Application.pdf	219046	yes	38
			157291521fe19c37955a2732aebd1e044dcad01f		
	Multipart Description/PDF files in .zip description				
	Document Description		Start	End	
	Specification		1	30	
	Claims		31	37	
	Abstract		38	38	
Warnings:					
Information:					
3	Drawings-only black and white line drawings	IP-0645-PRV_Figures.pdf	310599	no	1
			6980c835c170939350d3727d987dd00a3e5ae811		
Warnings:					
Information:					
4	Fee Worksheet (SB06)	fee-info.pdf	29440	no	2
			e12887b89310a99a79a836c3d5847164dec28d64		
Warnings:					
Information:					
Total Files Size (in bytes):			2036400		

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

New Applications Under 35 U.S.C. 111

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

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

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

New International Application Filed with the USPTO as a Receiving Office

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



UNITED STATES PATENT AND TRADEMARK OFFICE

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

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY,DOCKET,NO	TOT CLAIMS	IND CLAIMS
62/014,538	06/19/2014		260	IP-0645-PRV-A		

CONFIRMATION NO. 3783

29389
ILLUMINA, INC.
LEGAL DEPARTMENT
5200 ILLUMINA WAY
SAN DIEGO, CA 92122

FILING RECEIPT



Date Mailed: 07/03/2014

Receipt is acknowledged of this provisional patent application. It will not be examined for patentability and will become abandoned not later than twelve months after its filing date. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF APPLICANT, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection. Please verify the accuracy of the data presented on this receipt. **If an error is noted on this Filing Receipt, please submit a written request for a Filing Receipt Correction. Please provide a copy of this Filing Receipt with the changes noted thereon. If you received a "Notice to File Missing Parts" for this application, please submit any corrections to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections**

Inventor(s)

Jian-Bing Fan, San Diego, CA;
Kun Zhang, San Diego, CA;

Applicant(s)

Jian-Bing Fan, San Diego, CA;
Kun Zhang, San Diego, CA;

Power of Attorney:

Brent Moore--55461

If Required, Foreign Filing License Granted: 07/02/2014

The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 62/014,538**

Projected Publication Date: None, application is not eligible for pre-grant publication

Non-Publication Request: No

Early Publication Request: No

Title

METHODS AND COMPOSITIONS FOR SINGLE CELL GENOMICS

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent

in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at <http://www.uspto.gov/web/offices/pac/doc/general/index.html>.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, <http://www.stopfakes.gov>. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign Assets Control, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop technology, manufacture products, deliver services, and grow your business, visit <http://www.SelectUSA.gov> or call +1-202-482-6800.