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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 September 2011 (22.09.2011)

PCT

(10) International Publication Number
WO 2011/116088 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)   *C12N 15/11* (2006.01)

(21) International Application Number:
PCT/US2011/028657

(22) International Filing Date:
16 March 2011 (16.03.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/314,491   16 March 2010 (16.03.2010)   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):* **HENGEN, Paul, Nevin** [US/US]; 2434 E. Mountain Street, Pasadena, CA 91104 (US). **GEISS, Gary, K.** [US/US]; 647 NW 78th Street, Seattle, WA 98117 (US). **ELLIOT, Nathan, A.** [US/US]; 1800 Eastlake Ave. E #101, Seattle, WA 98102 (US). **WEBSTER, Phillppa, Jane** [US/US]; 4317 Densmore Ave. N., Seattle, WA 98103 (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, 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 GENOMIC FEATURES
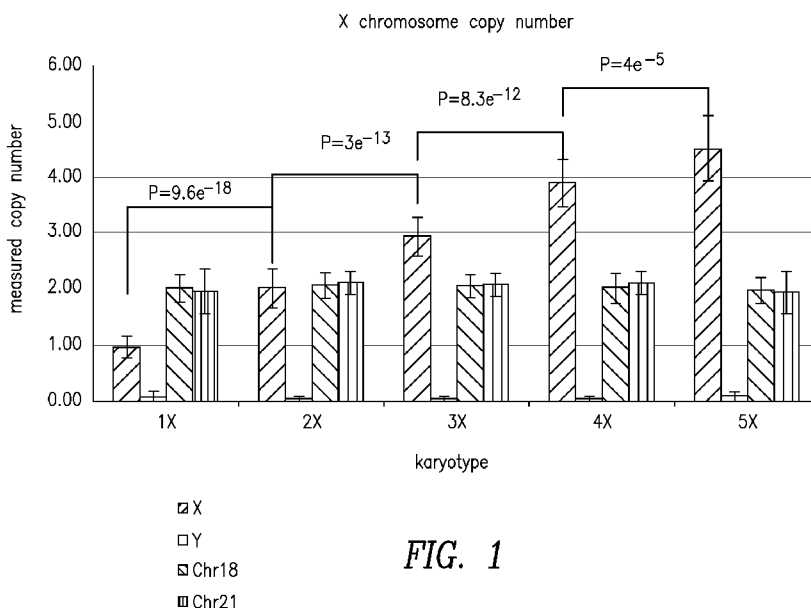
X chromosome copy number

FIG. 1

□X
□Y
□Chr18
□Chr21

(57) Abstract: The invention provides compositions and methods for the detection of gene copy number and/or chromosome copy number 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 2011/116088 A2

COMPOSITIONS AND METHODS FOR THE DETECTION OF GENOMIC FEATURES

RELATED APPLICATION

[01]   This application claims priority from U.S. Provisional Application No. 61/314,491, filed on March 16, 2010 which is incorporated by reference, 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]   Gene copy number and chromosomal number variations can have profound effects on biological functions. These variations are at the core of altered developmental, physiologic and pathologic processes. Therefore, identifying and quantifying the copy number of genes and chromosomes in subjects, in particular, in pre-natal subjects, can aid the early detection of pathology.

[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]   Thus, there exists a need for more accurate, quicker and more sensitive detection, identification and quantification of copy number of genes and chromosomes. Particularly, there exists a need for the specific detection of gene copy number and chromosome copy cumber in complex mixtures and multiplex reactions.

1

SUMMARY OF THE INVENTION

[07] The invention provides a method of detecting a target DNA sequence in a genome comprising providing a first sample containing genomic DNA; fragmenting the genomic DNA; denaturing the genomic DNA; providing a first nanoreporter comprising a first probe comprising a first label attachment region to which are attached one or more label monomers that emit light constituting a first signal; a second label attachment region, which is non-over-lapping with the first label attachment region, to which are attached one or more label monomers that emit light constituting a second signal; and a first target-specific sequence attached to the first · probe, wherein the target-specific sequence specifically hybridizes to the genomic DNA sequence to be detected; contacting the first probe with the fragmented genomic DNA wherein the contact is made under conditions sufficient for hybridization of the first target specific sequence to a fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected; stretching the first probe hybridized to the at least one fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected using a flow-stretch, receding meniscus, or electro-stretch technique, thereby spatially separating said label monomers, and measuring a signal from the first probe, wherein said signal uniquely identifies the at least one fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected.

[08] In one embodiment of this method of the invention, the first nanoreporter further comprises a second probe comprising a second target-specific sequence; and an affinity tag; wherein the first probe and the second probe specifically hybridize to the same fragment of the fragmented genomic DNA at different sites on the fragment. In another embodiment, the first probe further comprises an affinity tag.

[09] In another embodiment, the genomic DNA is mammalian genomic DNA. Specifically, the mammal is a human.

[10] In another embodiment, the genomic DNA sample is unamplified. In another embodiment, hybridization is performed in solution.

[11] In another embodiment, the fragmentation is performed by restriction enzyme digestion. Specifically, the restriction enzyme is Alu1 or Bfa1. In another embodiment, the fragmentation is performed chemically, enzymatically (for example, using one or more restriction endonucleases or a DNAase), by mechanical shearing or sonication.

2

[12]   In another embodiment, the signal generated from the first probe hybridized to the at least one fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected comprises about the same unit signal, or multiple thereof. In another embodiment, the signal generated from the first probe hybridized to the at least one fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected comprises a mixture of two or more different label monomers.

[13]   In another embodiment, said labels are fluorescent.

[14]   The invention also provides a method of determining the copy number of the target DNA sequence to be detected as described above further including the steps of providing a reference sample comprising fragmented genomic DNA wherein the copy number of the genomic sequence the first target specific sequence specifically hybridizes to in the reference sample is known; contacting the first probe with the reference sample wherein the contact is made under conditions sufficient for hybridization of the first target specific sequence to a fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected; stretching the first probe hybridized to the at least one fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected using a flow-stretch, receding meniscus, or electro-stretch technique, thereby spatially separating said label monomers; measuring a signal from the first probe, wherein said signal uniquely identifies the at least one fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected; and comparing the signal from the first sample to the signal from the reference sample, wherein the copy number of the first sample is determined by correlating the signal from the first sample with the signal from the reference sample.

[15]   The reference sample can be a synthetic nucleic acid sample or a biological genomic DNA sample.

[16]   The invention also provides a method of normalizing the signal generated as described above, further comprising providing at least a second nanoreporter comprising a third probe comprising a third label attachment region to which are attached one or more label monomers that emit light constituting a third signal; a fourth label attachment region, which is non-over-lapping with the third label attachment region, to which are attached one or more label monomers that emit light constituting a fourth signal; and a third target-specific sequence attached to the third

probe, wherein the target-specific sequence specifically hybridizes to a first DNA fragment from a copy number invariant region of the genome; contacting the third probe with the fragmented genomic DNA from the first sample and the reference sample wherein the contact is made under conditions sufficient for hybridization of the third target specific sequence to the first DNA fragment from a copy number invariant region of the genome; stretching the third probe hybridized to the first DNA fragment from a copy number invariant region of the genome using a flow-stretch, receding meniscus, or electro-stretch technique, thereby spatially separating said label monomers; measuring a signal from the third probe, wherein said signal uniquely identifies the first DNA fragment from a copy number invariant region of the genome; and comparing the signal from the the the second nanoreporter contacted with the first sample and the second nanoreporter contacted with the reference sample, wherein the number of multiples of the quantity of signal from the second nanoreporter contacted with the first sample compared to the quantity of signal from the second nanoreporter contacted with the reference sample normalizes the signal from the first nanoreporter contacted with the first sample.

[17]   In one embodiment of this method, the second nanoreporter further comprises a fourth probe comprising a fourth target-specific sequence; and an affinity tag; wherein the third probe and the fourth probe specifically hybridize to the same first DNA fragment from a copy number invariant region of the genome; at different sites on the fragment.

[18]   In another embodiment, the first DNA fragment from a copy number invariant region of the genome comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1-66. In another embodiment, the first DNA fragment from a copy number invariant region of the genome comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 2, 5, 7, 12, 13, 17, 19, 24, 25, 28, 32, 36, 38, 40, 44, 46, 50, 52, 56, 58, 62 and 66. In another embodiment, the first DNA fragment from a copy number invariant region of the genome comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 2, 5, 13, 19, 28, 46, 50, 56, 58 and 66.

[19]   In another embodiment, the first probe further comprises an affinity tag. In another embodiment, the genomic target DNA is mammalian genomic DNA. Specifically, the mammal is a human.

[20]   In another embodiment, the signal generated from the second probe

4

hybridized to the first control DNA sequence comprises a mixture of two or more different label monomers.

[21]   In another embodiment, the labels or label monomers are fluorescent.

[22]   The invention also provides a method of detecting a DNA sequence in a genome including the steps of   providing a first sample containing genomic DNA; fragmenting the genomic DNA; denaturing the genomic DNA; providing a first nanoreporter that specifically hybridizes to the genomic DNA sequence to be detected; contacting the first nanoreporter with the fragmented genomic DNA wherein the contact is made under conditions sufficient for hybridization of the first nanoreporter to the genomic DNA sequence to be detected; stretching the first nanoreporter hybridized to the at least one fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected using a flow-stretch, receding meniscus, or electro-stretch technique; and measuring a signal from the first nanoreporter, wherein said signal uniquely identifies the at least one fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected.

[23]   Various embodiments of this method include the following. The first nanoreporter is a single or a dual nanoreporter. The genomic DNA can be mammalian genomic DNA. Further, this mammal can be a human.

[24]   The genomic DNA sample can be unamplified. The contacting step can be performed in solution. The fragmentation can be performed by restriction enzyme digestion. The restriction enzyme can be, for example, Alu1 or Bfa1. The fragmentation can also be performed chemically, by mechanical shearing or sonication.

[25]   The invention also provides a method of determining the copy number of the target DNA sequence to be detected as described above including the steps of providing a reference sample comprising fragmented genomic DNA wherein the copy number of the genomic sequence the first nanoreporter specifically hybridizes to in the reference sample is known; contacting the first nanoreporter with the reference sample wherein the contact is made under conditions sufficient for hybridization of the first nanoreporter to a fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected; stretching the first nanoreporter hybridized to the at least one fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected using a flow-stretch, receding meniscus, or electro-stretch technique, measuring a signal from the first nanoreporter, wherein said signal

uniquely identifies the at least one fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected; and comparing the signal from the first sample to the signal from the reference sample, wherein the copy number of the first sample is determined by correlating the signal from the first sample with the signal from the reference sample.

[26]   The reference sample can be a synthetic nucleic acid sample or a biological genomic DNA sample.

[27]   The invention also provides a method of normalizing the signal generated as described above. This method further includes the steps of providing at least a second nanoreporter; contacting the second nanoreporter with the fragmented genomic DNA from the first sample and the reference sample wherein the contact is made under conditions sufficient for hybridization of the second nanoreporter to a first DNA fragment from a copy number invariant region of the genome; stretching the second nanoreporter hybridized to the first DNA fragment from a copy number invariant region of the genome using a flow-stretch, receding meniscus, or electro-stretch technique; measuring a signal from the second nanoreporter, wherein said signal uniquely identifies the first DNA fragment from a copy number invariant region of the genome; and comparing the signal from the the second nanoreporter contacted with the first sample and the second nanoreporter contacted with the reference sample, wherein the number of multiples of the quantity of signal from the second nanoreporter contacted with the first sample compared to the quantity of signal from the second nanoreporter contacted with the reference sample normalizes the signal from the first nanoreporter contacted with the first sample.

[28]   Various embodiments of this method include the following. The second nanoreporter can be a single or a dual nanoreporter.

[29]   The first DNA fragment from a copy number invariant region of the genome comprises a nucleic acid sequence can be any of SEQ ID NOs: 1-66. Preferably, the first DNA fragment from a copy number invariant region of the genome comprises a nucleic acid sequence can be any of SEQ ID NOs: 2, 5, 7, 12, 13, 17, 19, 24, 25, 28, 32, 36, 38, 40, 44, 46, 50, 52, 56, 58, 62 and 66. More preferably, the first DNA fragment from a copy number invariant region of the genome comprises a nucleic acid sequence can be any of SEQ ID NOs: 2, 5, 13, 19, 28, 46, 50, 56, 58 and 66.

[30]   The genomic DNA can be mammalian genomic DNA. Further, this mammal can be a human.

[31]   The signal generated from the first or second nanoreporter hybridized to the at least one fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected can include a mixture of two or more different label monomers.

[32]   Further, the labels or label monomers can be fluorescent.

[33]   The invention also provides a method of detecting two or more DNA sequences in a genome including the steps of providing a first sample containing genomic DNA; fragmenting the genomic DNA; denaturing the genomic DNA; providing a two or more nanoreporters that each specifically hybridize to a distinct genomic DNA sequence to be detected; contacting the two or more nanoreporters with the fragmented genomic DNA wherein the contact is made under conditions sufficient for hybridization of the two or more nanoreporters to the genomic DNA sequence to be detected; stretching the two or more nanoreporters hybridized to its corresponding distinct genomic DNA sequence using a flow-stretch, receding meniscus, or electro-stretch technique, and measuring a signal from the two or more nanoreporters, wherein said signal uniquely identifies each of the corresponding distinct genomic DNA sequences thereby detecting two or more DNA sequences in a genome.

[34]   Various embodiments of this method include the following. The two or more nanoreporters can be single or dual nanoreporters.

[35]   The genomic DNA is mammalian genomic DNA. Further, the mammal can be a human.

[36]   The genomic DNA sample can be unamplified. The contacting step can be performed in solution.

[37]   The fragmentation can be performed by restriction enzyme digestion. Restriction enzymes to be used include Alu1 and Bfa1. The fragmentation can also be performed chemically, by mechanical shearing or sonication.

[38]   The invention also provides a method of determining the copy number of the target DNA sequence to be detected as described above further including the steps of providing a reference sample comprising fragmented genomic DNA wherein the copy number of the genomic sequence the first target specific sequence specifically hybridizes to in the reference sample is known; contacting the first nanoreporter with the reference sample wherein the contact is made under conditions sufficient for hybridization of the two or more nanoreporters each fragment of the fragmented

genomic DNA comprising the genomic DNA sequences to be detected; stretching the two or more nanoreporters hybridized to each fragment of the fragmented genomic DNA comprising the genomic DNA sequences to be detected using a flow-stretch, receding meniscus, or electro-stretch technique, measuring signals from the two or more nanoreporters, wherein said signal uniquely identifies each of the fragments of the fragmented genomic DNA comprising the genomic DNA sequences to be detected; comparing the signal from the first sample to the signal from the reference sample, wherein the copy number of the first sample is determined by correlating the signal from the first sample with the signal from the reference sample.

[39]   The reference sample can be a synthetic nucleic acid sample or a biological genomic DNA sample.

[40]   The invention also provides a method of normalizing the signal generated as described above. The method includes the steps of providing at least one invariable sequence specific nanoreporter; contacting the at least one copy number invariant sequence specific nanoreporter with the fragmented genomic DNA from the first sample and the reference sample wherein the contact is made under conditions sufficient for hybridization of the at least one invariable sequence specific nanoreporter to a first DNA fragment from a copy number invariant region of the genome; stretching the second nanoreporter hybridized to the the first DNA fragment from a copy number invariant region of the genome using a flow-stretch, receding meniscus, or electro-stretch technique; measuring a signal from the second nanoreporter, wherein said signal uniquely identifies the first DNA fragment from a copy number invariant region of the genome; and comparing the signal from the the second nanoreporter contacted with the first sample and the second nanoreporter contacted with the reference sample, wherein the number of multiples of the quantity of signal from the second nanoreporter contacted with the first sample compared to the quantity of signal from the second nanoreporter contacted with the reference sample normalizes the signal from the first nanoreporter contacted with the first sample.

[41]   Various embodiments of this method include the following. The copy number invariant sequence specific nanoreporter can be a single or dual nanoreporter.

[42]   The first DNA fragment from a copy number invariant region of the genome comprises a nucleic acid sequence can be any of SEQ ID NOs: 1-66. Preferably, the first DNA fragment from a copy number invariant region of the genome comprises a

nucleic acid sequence can be any of SEQ ID NOs: 2, 5, 7, 12, 13, 17, 19, 24, 25, 28, 32, 36, 38, 40, 44, 46, 50, 52, 56, 58, 62 and 66. More preferably, the first DNA fragment from a copy number invariant region of the genome comprises a nucleic acid sequence can be any of SEQ ID NOs: 2, 5, 13, 19, 28, 46, 50, 56, 58 and 66.

[43]   The genomic DNA is mammalian genomic DNA. Further, the mammal can be a human.

[44]   The signal generated from the two or more nanoreporters hybridized to each of the fragmented genomic DNAs comprising the genomic DNA sequence to be detected can include a mixture of two or more different label monomers. The signal generated from the invariable sequence specific nanoreporter can include a mixture of two or more different label monomers.

[45]   Further, the labels or label monomers can be fluorescent.

[46]   In any of the above methods, more than one invariable sequence specific nanoreporter may be used. The invariable sequences these nanoreporters specifically hybridize to can include any of SEQ ID NOs:1-66. As few as 1 and as many as 100 invariable sequence specific nanoreporters may be used at once. These nanoreporters may all specifically bind different invariable sequences. In one assay, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 invariable sequence specific nanoreporters may be used.

[47]   The invention also provides a kit, comprising a first nanoreporter comprising a first probe comprising a first label attachment region to which are attached one or more label monomers that emit light constituting a first signal; a second label attachment region, which is non-over-lapping with the first label attachment region, to which are attached one or more label monomers that emit light constituting a second signal; and a first target-specific sequence attached to the first probe, wherein the target-specific sequence specifically hybridizes to a target DNA sequence; and a restriction enzyme.

[48]   In one embodiment of the kit, the first probe further comprises an affinity tag. In another embodiment, the first nanoreporter further comprises a second probe comprising a second target-specific sequence; and an affinity tag; wherein the first probe and the second probe specifically hybridize to the same target DNA sequence at different sites. Specifically, the restriction enzyme is Alu1 or Bfa1.

[49]   In another embodiment, the kit further comprises a second nanoreporter comprising a third probe comprising a third label attachment region to which are attached one or more label monomers that emit light constituting a third signal; a

fourth label attachment region, which is non-over-lapping with the third label attachment region, to which are attached one or more label monomers that emit light constituting a fourth signal; and a third target-specific sequence attached to the first probe, wherein the target-specific sequence specifically hybridizes to a control DNA sequence.

[50]   In another embodiment, the second nanoreporter further comprises a fourth probe comprising a fourth target-specific sequence; and an affinity tag; wherein the third probe and the fourth probe specifically hybridize to the same control DNA sequence at different sites on the sequence.

[51]   In another embodiment, the first probe further comprises an affinity tag. In another embodiment, said labels are fluorescent. In another embodiment, the kit further comprises a control DNA sequence of known copy number.

[52]   The invention also provides a composition comprising an isolated nucleic acid probe comprising at least 50 nucleotides of a sequence selected from the group consisting of SEQ ID NO: 1-66.

[53]   The invention also provides a method of selecting probe pairs for detection of a genomic sequence comprising providing the genomic sequence; performing *in silico* restriction fragmentation of the genomic sequence; generating *in silico* probe pairs for every position on the *in silico* restriction fragments, wherein each member of each pair is 35-50 nucleotides in length, and wherein each member of each set of pairs is complementary to a contiguous sequence; discarding sets of probe pairs wherein the melting temperatures of the probe pairs differ by more than 5° C; subjecting the remaining probe pairs to BLAT scoring; and discarding sets of probe pairs with the lowest 75% of BLAT scores; thereby selecting probe pairs for detection of a genomic sequence.

[54]   In one embodiment of this method of selecting probe pairs further includes scoring the fitness of the probes based on their length. Preferably, the fitness scoring is based on (i) the length of the restriction fragment the probe sequence optimized for use with the nCounter® system (also referred to as the nanoreporter code system). It can also be based on (ii) the location of the probe within the region to be analyzed. The location score is based on the number of probe pairs needed per region when the assay is designed. It can also be based on (iii) the results of the BLAT scoring, which provides a measure of whether the Reporter probe can be unambiguously mapped within the same reference genome sequence.

[55]   Further, any of the nanoreporters described above may be detected using the nCounter® system (alo referred to as the nanoreporter code system).

BRIEF DESCRIPTION OF THE DRAWINGS

[56]   Figure 1 is a bar graph depicting detection of X chromosome copy number in comparison to Y chromosome and chromosome 18 and 21 copy number using the nCounter® system of the invention.

[57]   Figure 2 is a bar graph depicting detection of X chromosome, Y chromosome and chromosome 18 and 21 copy number using the nCounter® system of the invention.

[58]   Figure 3 is a bar graph depicting detection of X chromosome copy number using the nCounter® system of the invention.

[59]   Figure 4 is a bar graph depicting detection of kinase and GPCR DNA copy number using the nCounter® system of the invention. These data demonstrate the ability of the nCounter® system to measure fold-changes and detect 2 copies per cell with negligible error.

[60]   Figure 5 is a bar graph depicting detection of kinase and GPCR RNA molecules using the nCounter® system of the invention. These data demonstrate the ability of the nCounter® system to measure small fold-changes in other nucleic acids that are similar to DNA copy number changes expected in the CNV assay.

[61]   Figure 6 shows 3 bar graphs showing copy number of genomic sequences in comparison to 23 invariant genomic regions in three different samples using the nCounter® system of the invention.

[62]   Figure 7 is a line graph showing the relationship between fragment length and counts obtained via hybridization using the nCounter® system.

[63]   Figure 8 is a bar graph showing a comparison of copy number for a region of chromosome 7.

[64]   Figure 9 is a chart showing the number of copy calls for all 313 autosomal probes in the Human Karyotype Panel.

DETAILED DESCRIPTION

[65]   The invention provides a sensitive, hybridization-based technology for determining the copy number of a given DNA sequence in a genome. For instance, the copy numbers of genes, intragenic sequences, intronic sequences, regulatory

sequences (including promotor, enhancer, and repressor elements), and gene splice forms can be determined. In certain embodiments of the invention, the nCounter® Analysis System is used. In preferred embodiments, the genomic DNA to be used with the nCounter® Analysis System is fragmented.

[66]   In specific embodiments, genomic DNA is fragmented into sequences of between 100 and 5000 bp prior to use with the nCounter® Analysis System. More specifically, the fragments are on average between 100 and 1000 basepairs ("bp"), between 100 and 500 bp or between 200 and 500 bp. According to certain embodiments, the average DNA fragment size of a sample of genomic DNA for use with the nCounter® Analysis System according to the invention is within one standard deviation of 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900 or 2000 bp.

[67]   In specific embodiments of the invention, fragmentation of genomic DNA is performed using enzymatic digestion, sonication, mechanical shearing or chemically. In a preferred embodiment, enzymatic digestion is performed using restriction enzymes. Preferred restriction enzymes include Alu1 which cuts as sequence AGCT. Another preferred restriction enzyme is Bfa1 which cuts at sequence CTAG. More than one restriction enzyme may be used to fragment the genomic DNA. In specific embodiments, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 restriction enzymes are used to digest the genomic DNA.

[68]   According to certain embodiments, the genomic DNA must also be denatured from double stranded DNA to single stranded DNA. This denaturation may be done according to any method known in the art. In specific examples, the DNA is heat denatured. For example heat denaturation can be performed at 95°C for 5 minutes.

[69]   In specific embodiments of the invention, the copy number of a genomic sequence is ascertained. First, the signal from each sample is normalized by determining the relative amount of genomic DNA in each assay. This is accomplished by including a set of probes in each assay which determine the signal generated by invariant regions in the genome. "Invariant regions" are those regions of the genome whose copy number rarely differs between individuals. In a preferred embodiment, invariant regions with a copy number of two are selected. Probes to 1-9 invariant regions can be used. In another preferred embodiment, probes to ten such invariant regions are included. The amount of input DNA in each assay can be deduced from the relative signal generated by these invariant regions in each sample,

and the signal from all probes in each assay can subsequently be normalized to a common amount of input DNA.

[70]   Following this sample input normalization, the copy number of any given region in an experimental sample can be determined by comparing the signal of a particular probe in the experimental sample to the signal of the same probe in a reference sample in which the copy number of the genomic target of that probe is known. The reference sample can be either a synthetic sample containing known amounts of the appropriate targets, or, in a preferred embodiment, a biological genomic DNA sample in which the copy number of the relevant regions is known. nCounter®

[71]   Probes to any of the following sequences can be used as control sequences.

Table 1

SEQ ID NO:	Chr	Start	End	Sequence
1	chr1	74174289	74174389	TCCCTATTACTCACCTTTCCTTTATAGTACAAGAGTTGTGTGTACTCACTCTCCCTGTG
2	chr1	97239054	97239154	TTACTTTCCAGCAATTATCCCTTGAACTCCTTTAACAAAGCT TAACATTTGATGTGTACCGACTCCAGGTAAGTTGCTGCTGTGTTCACATTTAACGTGTTAT
3	chr1	145108662	145108762	CTCATTCAGTCCATCTGATTACCCCATGAGTTGTTGCTATCT GTGAGAGTGATTAAAAGGAGCACTTATTGGGGTTTATTCTCCAATTCTCCATTCTTAT
4	chr2	83363170	83363270	TTGGGCTATCTACAAGCCATCTGAATTTACTTCTTCTTCAAAGCT ATCTTGATAAGTCATCTATCATTTGGCCAGCAACCTTTGGAATTCGAGGAACAGCAAC
5	chr2	137221997	137222092	GGTGATTGATGTGTTGCTAAATCGTGAAAATTGCTTCATACA GAGTTGACCACAAAGTGCTTTAGAACCAACCCTCTGATGCATAAGCCATCACACAATT
6	chr2	170396719	170396819	TTGCGTTTGCAGATTTAAGACAATACTTTCATCAAAA GATTCAAGATTCTTCTCCATCTCCAATCAAGAACATCCTACAGGCAACAGTCAGACAC
7	chr3	65580558	65580653	TGGTGAAAAGGATTATGGCACAGAAAGATTTATTCTCATTTA CCCTTGGCATTGATGTTTTATGGCTGTGTGAAACCAGTAAATCAGGAATTTGGAAAGCAG
8	chr3	123885091	123885186	AAATGGTGTTCCTACTAAAAGAGAAAAACAGGTAACAA CAAGTGGACAGTCTTCCACCTGCACCCAACCTTTAGATCATCAACATTTAATTTTATTTA
9	chr3	182367754	182367854	AGGAGAAGAGATGGGTTCTTACCACAAGGAACCCGACAA TGATTTTACCTCCCAAATACATAATCGTCTACTTCTCTTGTATGCTTCTGGGGACCCT
10	chr4	87783205	87783305	TGTCCATACATTATTATCTATCTCGGAACTGGTGCAATAGCT TGACTACTTATAAAGGGATTTATATCCTTGCCTAAGTGAGGGAAATTGAGTCTCTTCC
11	chr4	126464576	126464676	TCTGACTAATCAATCAGTCTAACACCCTAAAACCTCAAGATT ATAACAAGTAGCGTAAGTAAAGGTGTATAACTGGACACATCTTTGGAATGAACAATAT
12	chr4	150216450	150216545	CAGAGTCATAGTCATTAGCCTCTTACCCCAAATGTCAGAGCT AAGTGGTATCAGTTAAGGTATAAAGACGTTGGATTTGGAATGTATCTGGAAGGTGGAG
13	chr5	71261235	71261335	TTCATGGTATTCGATGGATGAATAGGCTATGGGGCATA ATAAACCCTCAGATTCCCGAATCTCCATATATCACATCCTCAACCGGGTAGTGGTGTTT
14	chr5	89956484	89956584	GATCATCTTCTCCAGATATACAGAAGAATCACAGGAGATCTG ACTAACCCATACATCATTCTGCCCTTCAAATTAATAGGTCCTATAACGTAATTATAGA
15	chr5	92902783	92902883	TATTGACTATAAACTGCATTAAAACTACGTGGTTTCTACGAGCT TTATTGATTTCCATTGAGGGAGGACTGAAAGACCATGGCTTAACATACTGTTGTTCCCA GGAGGAAAAAGTATAGTGAGAACTCATTTATTCTGACTGGAA

16	chr6	11423907	11424007	GCTCTTTTCATCTCCACTCTCTTTATTGAGTCTTGCCCATTTACAACACAAAGGAGAGACT TGTTTTTGTAGATCTTTAAAACACACCTGAGGATTTTGAATTC
17	chr6	41868334	41868434	GTCTCTTGTGTGTGAATGCAAATTATTCTTACCAGCAAGTTCTGAAATTATGTTACAA TGTCCAAATTCTTTGTATTTGTGACTGTTCATCAAATGGCTG
18	chr6	158141177	158141272	ATGGTTGCTTACAAAGTCCCAAAGGCAGAGTGATAGAGGTAGGTGTTTGGCTACTCA AATACCATGGACAACACATTGCTATAGGTGATCTTATTT
19	chr7	41541152	41541252	CGAAATTATAAATTTTGGTGTTCTGGACAGAGTTACCACCTCCCTGATTTTTAGTGGA GCCCTCTTAAAATAAAGCATAAATGCCAACAGCAACCAAATCA
20	chr7	68962045	68962145	TAAGTGAGGCTAAAGCATATCCCACTGGAGTAATTTTAAACCCTCTTTGCTTCCTTT GAAAGCACAATATCCTCAAATCTTATACAGTGGTTAATAGCT
21	chr7	84761978	84762076	TCAAAGATGTTTGTGTCCCAATCCCTTGAACATGAGAATAGTTCTATTAAATGGCAAC AGGGTGTTAAATTGTAGACGGAATTAAGGCTGTCAATCAG
22	chr8	32505847	32505947	ATTCAAATTGAGTCCAGGACTGGAATAGTCATGCAATTCTCCACAATAAAGATCTCAG TGAAACTTTATACAAGAACCAGTCGTGTGTTGAATTAAAGCT
23	chr8	77019431	77019528	CCTAATGAAGTATCGATCGGCTGCAGTTGAGGTAAAACAAGTTGACATTCATTTAGAG GTAGCAATGTGAATATACCACACAATGCCACCAAAGAAATA
24	chr8	93101553	93101653	ACTTGAAACACACATTTTCCGCATTATCTTAAATCTCGTGTCTCCAGAGAGTGTTCATG TGTATGTGGATTAGGCCTGGCTGGTTATTTCACCATCTATCT
25	chr9	72376995	72377095	ATATAGGTGGTTATTTAAGACTTTAGCATTAAGCGGAGACTGGGTAAAGAATATACAGA ACAGGAATTTGTGTACTAACTCTGCAACCTTTCTACAAATC
26	chr9	85571439	85571539	TCAGTACTATGTAACTTTTGCCATAGTCTCATTTCTACTTAACAGTCCTTTGATTGAT CTCTTTCTAAATTTGACAGATTCCTACAATAGCCATTCCAAA
27	chr9	139587089	139587184	GCTTTTTGCCTGATACCTGTTGGATACAATAGTGACTTCGCCATGTTGGATTTCAAGTT TGTCTGCACCCTTTTTGGATATAGTTTATCTTCCAGA
28	chr10	76411621	76411718	GAATTTTTTATTTACACGCGCTTACCTCAAAATACTGAAAAGGTCTTTCCTTCGGTAAA TTTCATTTGCTGGAGGATGAAACCATCCACACTTCTTGG
29	chr10	97648682	97648782	TTCATTTTTCATCATAGACACAGTTAATCACAGAGACTTACTCCAACCCAGAAGTTTATACC AAAAGACTTGATTAGCGCAATAAAAGCACTAAATGAGGAGCT
30	chr10	112756032	112756132	ACAACAACAACAAAAAAAAAATTTGCTGCACATGGATTGTCTCATTCCACACAA AATGACCCATAAAAGAAATCCCAAGTAGCAAGCATCCATTTG
31	chr11	62627341	62627441	TAGGTGCTATCATCATAGGTTATATGCAGAGTATTTTTGAGAAAAACACAGAGGAAGTATC CACTTTAGCCTTAATAGGTCAGAAAAAGGATTAGGAGTTAGCT
32	chr11	93785528	93785628	CACCCAGTTAGACTATTCGACACAAATCATCATACGATATTATAAAAGGCTGAAATTAAGGGT

AGAGTGATAAAAATCGAAATTGTGTGAAGAAAATGACCATGT

No.	Chr	Pos 1	Pos 2	Sequence
33	chr11	120914392	120914492	CCTGTTGATTGATTGTATAATAAGATCCATAAGAGAAAGAAGGATCTCAGGTATTTTAGTTAAAGTGAACTCAGCCTACTGATACCAGTTAAAAGATT
34	chr12	19013949	19014049	CAATATCGATTCTCTACGTCTCTCAGAAATTGGTTGCCTAACAACTTTTTGCTCAATAAATTTTGGAGACTCCTGGGATTGGTGCTTATCGAGAAACATA
35	chr12	45584004	45584104	TTTTCTCTACTGAAACTTGTTCTCTCCTAAAAATATACGCCAGTTGCTAAGTATTCAGCATTGACTTTTCTACCACAGAATACCCATAACAAC
36	chr12	87562270	87562370	GATTCAGATCTCCTCTTTTAAGATGTGATGGCCTCATTCCACTAAGTATGTAAACCAAACCTTTACCAAAGCACCAGGCATTTGATTAAAGATTCACTA
37	chr13	39406521	39406621	ATCTGAAAATGTCTCCAGATAAGTCTATTGTGAATCACTTTGCATTAATTATATACCCAGTAACAGATTAAGTCCATCCAATTTGAAGACCCACATCTTAC
38	chr13	45678720	45678820	ACAACATTAGAAGGGATTGCTTCCAGAGATTGTAACTTGGTGTATCACTTTACCCAAGTGTTCCTACTTAAGAAAGAAAAAGCAAAGTGCCTCACCC
39	chr13	100916026	100916126	CAGGGATCCTCAACCTCATACCTTCTCTTCAAAAAAGTCAGAAGTACCATACCAAATATAAATGGGTGACTGTTATTGCCAAGATCACATAGTAGATAA
40	chr14	55784606	55784706	GACATGTTCGTTGCATAATAGCAGCAGCATGGTAGACGCTGAAAATTATTTTGGACTGTATTTCACATTTAGGCAACTACTTTAATGGTTAAATCAACCC
41	chr14	66138914	66139014	AAAATTCGTATTCACATTTCAAGTTATATGTGTCAAAGCACTGGTGCTGAAACAGAATAGGTTATCTTCTAATTTCACATCACTGAGTTATTCACTGCAG
42	chr14	88109044	88109144	GATTTTTAGCCTAAGCCAGAATTTAAAGTACATACAAACCTCCATACTCATTTTCTCCGAGTTGTTTCTAAAGAACGGACTATACGTTTCTTCTAAGCT
43	chr15	32163647	32163747	ACTACTATAAACTTGAGTCATCCCGACGTTGATCTCTTACAACTGTGTATGTTAACTTTTTAGCACATGTTTGTACTTGGTACACGAGAAAAACCCAGCT
44	chr15	35899734	35899833	CTTACTAAATAGTGGAATGGAGGGATAGTGAGCAACAACCTTGGAGCCAGAAGATGTAGTAATGAGACTCTGCTTTTGTCATTCACAGTATCTGTCAGCT
45	chr15	67068813	67068910	GCAAACTTACCTAATAATGGCTGTATGTATCATTACTTTCTGGAGTTCCTCTTATTTTGATGGGAACTTGCCTGCTTGGCTAAAACAGAAATGGCA
46	chr16	11026536	11026636	ATTTGACTGATTTCAGTTCTGATGTTAGGAAAGAGGTCAGACGCTAAGTCAGTTGTAAATCAAGGGGTCAAAAGAAAAACCACAGGGTGAATATAGTCATC
47	chr16	51412454	51412554	TGGCAAAATGCTGTTTTTCTATCAGTTGAAACATCTTTACCTTATTTGAGCTGTAGTGAAAGGGAGTTAAAAACATCTTTATAGTTGGGCCATAG
48	chr16	60328302	60328402	TCTATCATATGTGAAAACCGCCTGACTTTTGTGTTTAGCAAATTAATCTCAGACCACTTGTCACATCTTCTTGTCTTATCAGCAATTGATATGGGCTTTTCCTTC

49	chr17	57079028	57079126	ACACATTTGATAAACTTTTATCTTAATGCGCCTTTCTGGAATACCAGTCTGACCTCAA TCTGAAACAAAGCCTTAGTTGATGATGTTTGCAGGAGGTAG
50	chr17	60617471	60617571	AAACATATTGAAGGAAGGCACTAAACAAAACAGCATCTTCAGTCCCGATTAGTACCAT GACTTGAGTCTTACACAGTCAGAATACATGATTAGTCACATC
51	chr17	77035037	77035137	TTTGTGACATGAGAAGCCCTGAGATTAATTTTTGCCTGTCTTAATTGAAGGAACCATTT AGTGCCGATTAACTATTATTACCAAATCATCAGGATTGATG
52	chr18	20244655	20244755	ATAATTCCTGAGAATGTGTTTATGTGCTGTGTGGTGATACGTCAGTTGCATCCTCTCCTTT ATACCCACACATTGACTAAGTCACAGTACCTTATGTTCTTCA
53	chr18	30352718	30352818	GAAAATATTGCTATATGTACCTCCCCCACTATACCAGGAGATATTTCAGGTGCTGCAT TCTATTAATGTTCCCGTCTTTACTACCTAATAGTGTCTCACA
54	chr18	71084633	71084733	TAGTACATAAAAAAATGTTGGACTCTCAGGCTAATTTAGGGTTGCTAAGTCAAAAGAT TGATGTTACAGGTGAAAAATACATAGTGTGCCTGTCATTCTCCTA
55	chr19	38623780	38623871	CAGATGCCATAGGTGGGGCCAGAACCATCTAAAACATTACCTGTAGGGTTGTTCCATTTC AGACAACTCCAATTGACCATTCAGAGGGTTTG
56	chr19	38818849	38818949	GCCGCATCAAATTAGCATCGACTCGTAAAACGTTACTGAATGATTCCTCAAATCTGCC AAGTCTTCAGATCAATTTTGGAGAAAGCGTCAAGAGGTTTTT
57	chr19	53490523	53490623	GGTGTAGGAGTGAGAGGGCTTAGAGAACACCTTGATAACTCTTTCCTGTAGTTGAGTCAT GCCAAATGCCCTGTCAAAATTTAATCCATTGGTATCAAAGCT
58	chr20	10504897	10504992	GCCCAGGGATTCTTAATGCTTCACAAATAAGTCACTCTGAATCTGTGGCAAATT CACTTAGAGACAGTATAAAATGTCTATCGTACCAAAGG
59	chr20	18616057	18616157	ACTATAAATACCTCCTTTTACTTCCTACAGTTCACTAAGTCTAACCTGGGCTACCACT GTGGAAGAGAGATTTCTCCTTTATCAGAAGGCACTTCAGCAAAC
60	chr20	49452575	49452675	CTGACCTGCTTACAATTCCATCTCTCTTGGATAAGCAAAGAGAGGCATATACTCAAATGT CTTAAAAGAAATGTTTGGTTAATTCCTCTAACCCCCAGAGGG
61	chr21	32806387	32806487	CATTCTCTTCATTGGTCAATACATAGCCCTACTTTATGTCTAACGAATTACTTTTTAA TACTGTAATTAGCACCAGTGCTATGAATGCACACCCGTATAG
62	chr21	36113846	36113941	CAAATAACAAAACTCAGAAAGGCTGTTGTCAATGTAAAACTTGACTCCTAAGCAAGGA TTCCCTTGTTGAATACAAAGTAAAGAAGAGCAGCACTGG
63	chr21	37658395	37658495	GACCTGGTTCACAGATGAAATCCTTGTCATCTAAGAATCTTCCCATTAGATTCACTTA CAGATGTGTTTATTCATAGACTGTTCACCTTGAAAAGCAAAG
64	chr22	26799325	26799423	CTTCTCATCTTCCTTTTGCTCCAAAACTATGGGCACTCCTTGGTTAATGGACATTCCTT TAGAAATTTGATCTATCCCAAGGACACACAGATATATGTCCC
65	chr22	39576282	39576382	GCTCTACTACCTGAGTGATATTTGTGAGTGTGAATCATGGTGTTGGGTTAGCATATTT

66	chr22	42509371	42509471	GCTTAAAGGACGTGTAAGATTAGGAGAAGGTTACCAGTAGCTACACATTCAAGACCCCATTCTTCACCGTGTAGAGTATATTCAAGGAATGGTTCCCCAAATAAGTTCAGATCTTCTTCAAGTAAGTATTCATGAGCAAATA

[72]   The sequences shown in Table 1 are sequences that can be used as invariant controls for each of the non-sex chromosomes. Of the nucleic acid sequences shown above in Table 1, SEQ ID NOs: 2, 5, 7, 12, 13, 17, 19, 24, 25, 28, 32, 36, 38, 40, 44, 46, 50, 52, 56, 58, 62 and 66 are the preferred sequence for each corresponding chromosome. Of these sequences, SEQ ID NOs: 2, 5, 13, 19, 28, 46, 50, 56, 58 and 66 are particularly preferred for use as invariant controls for use in sample input normalization.

[73]   Nanoreporters can be used to detect any of the sequences of SEQ ID NOs: 1-66. Nanoreporters can also be used to detect fragments of these sequences. Fragments can be between 50 and 90 nucleotides in length. Fragments can also be 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98 or 99 nucleotides in length. The target specific regions of the nanoreporters can specifically hybridize to any of the sequences of SEQ ID NOs:1-66, any fragments thereof or any complement thereof.

[74]   The invention also provides methods of detecting the copy number of multiple genomic sequences simultaneously. Multiple probes made according to the nCounter® system of the invention can be used to detect multiple genomic targets in a multiplexed format. The signal generated by the presence of a number of copies of these targets can then be compared to a standard or reference sample with a known copy number. As explained above, a control or multiple controls are used to normalize sample DNA input between different assays.

[75]   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 a number of positions (visualized as "spots") and a number of colors. Colors are chosen to minimize spectral overlap during imaging. The number of positions are chosen with respect to a combination of factors that include, but are not limited to, 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 (for instance, in a

system with 7 positions, and 4 colors, the maximal number of specific labels is 4^7, or 16,384 labels, if all possible combinations of codes are used).

[76]   Specific capture and reporter probes are synthesized in 96-well plates using a semi-automated process. Briefly, the reporter probes are pooled. Gene-specific probes are ligated to reporter molecule backbones and the ligated backbones were annealed to a unique pool of dye-coupled RNA or DNA segments corresponding to a single code. Reporter probes are then purified using a common 5'-repeat sequence at the end of each backbone to remove excess probe oligonucleotides and dye-coupled RNA or DNA segments. Capture probes were made by ligating a second sequence-specific oligonucleotide for each gene to a universal sequence containing biotin. 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.

[77]   The expression levels of all DNA:reporter molecule:capture probe duplexes 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 gene-specific probes, as they were ultimately removed, and, thus, do not interfere with binding and imaging of the sample. All post hybridization steps are handled robotically on a custom liquid-handling robot (Prep Station, NanoString Technologies).

[78]   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 Technologies). The expression level of a gene is measured by imaging each sample in, for instance, 4 colors and counting the number of times the code for that gene is detected. For each sample, over 600 fields-of-view (FOV) are imaged (1376 X 1024 pixels) representing approximately 10 mm^2 of the binding surface.

Typical imaging density is 100-200 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- to 10-fold higher. The Digital Analyzer can accommodate up to 6 cartridges at once and current scan times for 600 FOV are 4 hours per sample cartridge. Unattended, the system can process 72 samples in 24-hours per instrument.

[79]   Image processing and code counting is performed. 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 do not meet all of these criteria are discarded. Using these criteria, approximately 40% of the detected molecules are typically counted. No parity schemes or error correction are employed in the current system. Data is output in simple spreadsheet format listing the number of counts per DNA, or genomic DNA fragment, per sample.

Therapeutic Applications

[80]   Compositions and methods of the invention are used to detect gene or chromosome copy number 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 or chromosome copy number 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 (onset of a genetic disease or degeneration of telomeres as a consequence of aging or increased cell proliferation due to cancer) or responses to genetic therapy. Furthermore, 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.

[81]   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. Furthermore, the present invention can be used to monitor the progression of a disease or responses to genetic 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 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.

[82]   Critically, the methods of the invention can determine the individual contributions of multiple genes to the same condition or disorder using one multiplexed reaction. Specifically, the methods of the invention can not only identify which genes are involved in a particular disorder, disease, or syndrome, but also by which mechanism, *e.g.* sequence duplication, mutation, deletion, or translocation. As such, the invention provides superior properties over all known methods of genetic screening because the methods provided elucidate the genes involved in complex multigene disorders such as Autism or Down Syndrome, which, unlike Huntington's Disease, require the participation of multiple genes. Most importantly, the methods of the invention do not require the skilled artisan to have identified or predetermined which genes may contribute to a particular disorder, disease, or syndrome prior to using these methods. If the skilled artisan applies gene specific probes that are specific for all known genes, then the abnormally increased or decreased genomic DNA copy numbers present in a sample taken from a subject having a particular disorder, disease, or syndrome will be apparent when compared to the genomic DNA copy numbers of a normal subject.

[83]   In certain embodiments of these methods, the term "normal subjects" is meant to describe a person who has not been diagnosed with the disorder, disease, or syndrome under examination. Alternatively, a "normal subject" is a person of similar age, weight, gender, ethnicity and physical health who has not been diagnosed with the particular disorder, disease, or syndrome under examination. In other aspects, a "normal subject" is a person who has not been diagnosed with any genetic disorder, disease, or syndrome,

and furthermore, may be of similar age, weight, gender, ethnicity and physical health to the test subject. In another aspect, a normal subject is a predetermined numerical reference based upon, for example, national or international averages or standards.

[84]   The methods of the invention encompass a variety of subjects. Subjects are plants or animals. Animals are 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.

[85]   The methods of the invention also encompass screening of subjects for pathology at various points in their lives. In certain embodiments, the detection is performed prenatally, neonatally, postnatally, at infancy, childhood, puberty, early adulthood, adulthood, and during old age. The pre-natal subject may be tested at about 1, 2, 3, 4, 5, 6, 7, 8 or 9 months prior to their expected birth date. Other subjects may be tested at the age of about 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, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, and 99 years of age or older.

Cancer

[86]   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.

[87]   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

23

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.

[88]   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.

[89]   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 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

[90]   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.

[91]   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 considered synonymous with mental 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).

[92]   A specific developmental disorder selectively affects one area of development, sparing essentially all other areas of development. Specific developmental disorders 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.

[93]   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, 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.

[94]   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.

[95]   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 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

[96]   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.

[97]   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; decreased movement; and ultimately, coma or death. Neurological disorders can result from genetic predisposition for developing the neurological disorder, one or more environmental factors that induce a the disorder to enhance the individual's genetic predisposition, or exposure of an individual to infectious agents such as a virus, a bacteria, a fungus, or a prion that induces the disorder or enhances the individual's genetic predisposition.

[98]   Exemplary neurological disorders include, but are not limited to, autism spectrum disorders (ASD), Angelman Syndrome, bipolar disorder, attention-deficit disorder (ADD), central auditory processing disorder (CAPD), cerebral palsy, Down Sydrome, expressive language disorder, IsoDendric 15 (abbreviated idic(15)), Lanau-Kleffner Syndrome, neural tube defects, seizure disorders, epilepsy, Tourette Syndrome, traumatic brain injury (TBI), childhood disintegrative disorder, agnosias (visual, auditory, and somatosensory), anorexia nervosa disorder, acute stress disorder, adjustment disorder, bipolar disorder, body dysmorphic disorder, breathing-related sleep disorders, brief psychotic episode, bulimia nervosa, schizophrenia, Huntington's Disease (HD), multiple sclerosis (MS), amyotrophic lateral sclerosis (ALS), Capgras (delusion) Syndrome, chronic motor or vocal tic disorder, circadian rhythm sleep disorder, cluttering (speech disfluency), conduct disorder, communication/language disorders, Creutzfeldt-Jakob

Disease (CJD), kuru, Gerstmann–Sträussler–Scheinker syndrome (GSS), fatal familial

insomnia (FFI), depression, addiction, Fahr's Syndrome (or idiopathic basal ganglia

calcification), migraine, neoplasm (benign and malignant), aphasia, paralysis, Bell's

Palsy, cerebrovascular disease, encephalitis, hydrocephalis, microcephaly, Parkinson's

Disease, trigeminal neuralgia, narcolepsy, muscular distrophy, Guillain-Barre Syndrome,

neurofibromatosis, dyslexia, Rett Syndrome, Fragile X syndrome,

adrenoleukodystrophy, ataxia telangiectasia, Cockayne syndrome, deafness, Duchenne

muscular dystrophy, Gaucher disease, Lesch-Nyhan syndrome, maple syrup urine

disease, Menkes syndrome, phenylketonuria, Prader-Willi syndrome, spinal muscular

atrophy, spinocerebellar ataxia, tuberous sclerosis, Neimann-Pick syndrome, Refsum

disease, Tay-Sachs disease, Charcot-Marie-Tooth syndrome, fibrodysplasia ossificans

progressive, myotonic dystrophy, and Meniere's Disease.

Stem Cell Disorders

[99]   Compositions and methods of the invention are used to identify cells and subjects

at risk of developing a "stem cell" disorder or those cells and subjects who may have a

predisposition for developing a stem cell disorder. Moreover, the compositions and

methods of the invention are used to differentiate stem cell disorders for the purpose of

diagnosing or prognosing a subject at risk of presenting or a subject who has been

diagnosed with a stem cell disorder. The compositions and methods of the invention are

further used to monitor to progression of a stem cell 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 stem cell disorder

himself/herself, or for producing a child having a stem cell disorder.

[100]   The term "stem cell disorder" includes any disorder that initially presents within

a totipotent (or omnipotent), pluripotent, multipotent, oligopotent, or unipotent stem cell

of an individual. Alternatively, or in addition, a stem cell disorder includes any disorder

which can be treated or prevented by administering a composition including a stem cell

to the individual. Stem cells are characterized by their ability to produce daughter cells,

one of which will differentiate and the other of which will remain an undifferentiated

stem cell. The potency of a stem cell relates to differentiation potential of the daughter

cell that becomes committed to a particular cell fate. Specifically, the terms totipotent

stem cell or omnipotent stem cell describe stem cells that can give rise to both embryonic

stem cells or, alternatively, the stem cell can generate every type of cell in the human body. Pluripotent stem cells have a more restricted potential than totipotent stem cells, however, these stem cells can generate cells derived from any of the three germ layers (ectoderm, mesoderm, or endoderm). Multipotent stem cells have a more restricted potential than pluripotent stem cells, however, these stem cells can generate cells within a related lineage. Multipotent stem cells are often considered adult stem cells because they are found in, for instance, the adult brain (neural stem cells that give rise to neurons and all types of glia) and bones (bone marrow stem cells that give rise to all types of blood cells). Oligopotent stem cells have a more restricted potential than multipotent stem cells, however, these stem cells can generate a few related types of cells. For example, the corneal epithelium contains oliopotent stem cells that produce only corneal and conjunctival cells. Unipotent cells are the most restricted cell type because they can only reproduce their own cell type, however, they do maintain the ability to self-renew. Muscle stem cells are nonlimiting examples of unipotent stem cells.

[101] Exemplary stem cell disorders include, but are not limited to, autism spectrum disorders (ASD), neural tube defects, seizure disorders, epilepsy, hearing loss, deafness, blindness, vision impairment, jaundice / kernicterus, cluttering (speech disfluency), agnosias (visual, auditory, and somatosensory), Huntington's Disease (HD), multiple sclerosis (MS), amyotrophic lateral sclerosis (ALS), chronic motor or vocal tic disorder, circadian rhythm sleep disorder, Alzheimer's Disease, dementia, senility, diabetes, Parkinson's Disease, muscular distrophy, Guillain-Barre Syndrome, sickle cell anemia or sickle cell disease, ataxia telagiectasia, Cockayne syndrome, DiGeorge syndrome, severe combined immunodeficiency, porphyria, paroxysmal nocturnal hemoglobinuria, thalassemia, familial Mediterranean fever, immunodeficiency with hyper- IgM, Charcot-Marie-Tooth syndrome, fibrodysplasia ossificans progressive, myotonic dystrophy, spinal muscular atrophy, spinocerebellar ataxia, and Gauchers Disease.

[102]

Nanoreporters and nCounter® System overview

[103] The basis of the nCounter® Analysis system is the unique code assigned to each nucleic acid target 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 DNA or RNA target, a biotinylated capture probe and a reporter probe carrying the fluorescent barcode. This system is also referred to, herein, as the nanoreporter code system.

[104] 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".

[105] The relative abundance of each target is 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 hybridization steps are handled robotically on a custom liquid-handling robot (Prep Station, NanoString Technologies).

[106] 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 sample input, and overall target abundance. Data is output in simple spreadsheet format listing the number of counts per target, per sample.

[107] 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 preferred nanoreporter is 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 nanoreporter probes is labeled. This labeled nanoreporter probe is referred to herein as a "reporter probe". The other nanoreporter probe is not necessarily labeled. Such unlabeled components of dual nanoreporters are referred to herein as "capture probes" and often 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. When both probes are labeled or both have affinity tags, the probe with more label monomer attachment regions is referred to as the reporter probe and the other probe in the pair is referred to as a capture probe.

[108] For both single and dual nanoreporters, 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 portion which provides a "code" of signals associated with the target-specific sequence. Upon binding of the nanoreporter probe to the target molecule, the code identifies the target molecule to which the nanoreporter is bound.

[109] 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 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.

[110] 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.

[111] 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.

[112] The nanoreporters described herein can comprise synthetic, designed sequences. In some embodiments, the sequences contain a fairly regularly-spaced pattern of a nucleotide (e.g. adenine) residue in the backbone. In some embodiments, a specific 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. 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 nucleotide that is not regularly-spaced but that is spaced at least an average of 8, 9, 10, 12, 15, 16,

34

20, 30, or 50 bases apart. 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 number of sites which can be attached to a detectable molecule. 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.

[113] In some embodiments, the nanoreporters described herein have a fairly consistent melting temperature (Tm). Without intending to be limited to any theory, the Tm of the nanoreporters described herein provides for strong 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 disclosure provides a population of nanoreporters in which the Tm among the nanoreporters in the population is fairly consistent. In some embodiments, the disclosure provides a population of nanoreporters in which the Tm of the complementary polynucleotide sequences when hybridized to its label attachment regions is about 80° C, 85° C, 90° C, 100° C or higher. In some embodiments, the disclosure provides a population of nanoreporters in which the Tm of the complementary polynucleotide sequences when hybridized to its label attachment regions is about 80° C or higher.

[114] 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° C, 1 X SSPE. 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.

[115] The disclosure also 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 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

36

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.

[116] In some embodiments, the sequence characteristics of the nanoreporter probes described herein provide sensitive 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.

[117] 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.

[118] The nucleic-acid based nanoreporter and nanoreporter-target complexes

described herein 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 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.

[119]   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. Publication No. 2010/0161026, which is incorporated by reference herein in its entirety.

[120]   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

38

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 diagnosis or prognosis of different biological states (*e.g.*, disease vs. healthy) by quantitative analysis of known biological markers or copy number variant loci.

[121] 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 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, uses one 6-position coded nanoreporter (also referred to herein as a reporter probe) and a capture probe. In some embodiments, a dual nanoreporter with a 6-position nanoreporter code can be used, using one capture probe with an affinity tag and one 6-position nanoreporter component. In some embodiments an affinity tag is optionally included and can be used to purify the nanoreporter or to immobilize the nanoreporter (or nanoreporter-target molecule complex) for the purpose of imaging.

[122] 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. 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.

[123] The disclosure 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 disclosure 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.*, 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 described herein, 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. Publication No. 2010/0112710, incorporated herein by reference).

[124] 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.

[125] 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.

[126] 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).

[127] The disclosure 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).

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

[129] In other embodiments, a nanoreporter includes spots providing like or unlike signals 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 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.

[130] In other embodiments, a nanoreporter includes 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.

[131] A detectable molecule providing a signal described herein 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.

[132] A nanoreporter can also be 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.

[133] In other embodiments, a nanoreporter also includes an affinity tag attached to the reporter probe 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.

[134] In other embodiments, a nanoreporter also includes 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 described herein 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.

[135] One embodiment of the disclosure provides increased flexibility in target molecule detection with label monomers described herein. 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.

[136] The disclosure 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.

[137] 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 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.

[138] 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

43

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.

[139] Additional disclosure regarding nanoreporters can be found in International Publication No. WO 07/076129 and WO 07/076132, the contents of which are incorporated herein in their entireties. Further, the term nucleic acid probes and nanoreporters can include the rationally designed (e.g. synthetic sequences) described in International Publication No. WO 2010/019826, incorporated herein by reference in its entirety.

Kits

[140] Kits include a composition containing at least a first probe and a restriction enzyme. The first probe includes a first molecule containing a first label attachment region to which are attached one or more label monomers that emit light constituting a first signal; a second label attachment region, which is non-over-lapping with the first label attachment region, to which are attached one or more label monomers that emit light constituting a second signal; and a first target-specific sequence attached to the first molecule, wherein the target-specific sequence specifically hybridizes to a target DNA sequence. In certain embodiments of the kit, the first probe further includes an affinity tag. In other embodiments of the kit, the first probe includes a second molecule containing a second target-specific sequence; and an affinity tag; wherein the first molecule and the second molecule specifically hybridize to the same target DNA sequence at different sites.

[141] In other embodiments of the kit, the first molecule is a reporter probe or a capture

probe. In other embodiments, the second molecule is a reporter probe or a capture probe. Reporter probes and capture probes are provided in the kit individually or in a mixture.

[142]   The restriction enzyme can be either a restriction endonuclease or a DNase. Preferably the restriction enzyme is Alu1 or Bfa1.

[143]   The kit optionally includes a second probe including a third molecule containing a third label attachment region to which are attached one or more label monomers that emit light constituting a third signal; a fourth label attachment region, which is non-over-lapping with the third label attachment region, to which are attached one or more label monomers that emit light constituting a fourth signal; and a third target-specific sequence attached to the third molecule, wherein the target-specific sequence specifically hybridizes to a control DNA sequence. In certain embodiments the second probe includes a fourth molecule containing a fourth target-specific sequence and an affinity tag; wherein the third molecule and the fourth molecule specifically hybridize to the same control DNA sequence at different sites on the sequence.

[144]   In other embodiments of the kit, the third molecule is a reporter probe or a capture probe. In other embodiments, the fourth molecule is a reporter probe or a capture probe. Reporter probes and capture probes are provided in the kit individually or in a mixture.

[145]   In preferred embodiments of the kits, the labels and label monomers are fluorescent.

[146]   Kits further include a control DNA sequence of known copy number. In other embodiments, kits contain a control probe to ensure that equivalent amounts of DNA are introduced or loaded into the nCounter® system.

[147]   Kits include instructions for handling the enclosed compositions and protocols for performing singular or multiplexed fragmenting and denaturing reactions, as well as the contacting, stretching, and measuring steps described herein using the enclosed compositions. Furthermore, the instructions provide guidance for preparing the resultant tagged fragmented and hybridized genomic DNA molecule(s) for detection using the nCounter® Analysis System.

Nucleic Acids

[148]   Also disclosed herein are isolated nucleic acid molecules that can be used as controls for genomic copy number assays. These nucleic acid molecules include the

45

sequences shown in Table 1. Of the nucleic acid sequences shown above in Table 1,
SEQ ID NOs: 2, 5, 7, 12, 13, 17, 19, 24, 25, 28, 32, 36, 38, 40, 44, 46, 50, 52,56, 58, 62
and 66 are the preferred sequence for each corresponding chromosome. Of these
sequences, SEQ ID NOs: 2, 5, 13, 19, 28, 46, 50, 56, 58 and 66 are particularly preferred
for use as controls.

[149] Also included in the invention are nucleic acid fragments sufficient for use as
controls in genomic copy number assays. These sequences can be present in the
genomic sample itself or can be added to separate control samples. Probes can also be
made to detect these sequences. These probes, themselves can be nucleic acid molecules
comprising any of SEQ ID NOs: 1-66 or any complement thereof. As used herein, the
term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or
genomic DNA), RNA molecules (e.g., mRNA), analogs of the DNA or RNA generated
using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic
acid molecule may be single-stranded or double-stranded, but preferably is comprised
double-stranded DNA.

[150] The term "probes", as utilized herein, refers to nucleic acid sequences of variable
length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as
approximately, e.g., 6,000 nt, depending upon the specific use. Probes are used in the
detection of identical, similar, or complementary nucleic acid sequences. In specific
embodiments, these probe sequences are used to construct target specific portions of
nanoreporters described above.

[151] The term "isolated" nucleic acid molecule, as utilized herein, is one, which is
separated from other nucleic acid molecules which are present in the natural source of
the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which
naturally flank the nucleic acid (i.e., sequences located at the 5'- and 3'-termini of the
nucleic acid) in the genomic DNA of the organism from which the nucleic acid is
derived. For example, in various embodiments, the isolated nucleic acid molecules
described herein can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb
of nucleotide sequences which naturally flank the nucleic acid molecule in genomic
DNA of the cell/tissue from which the nucleic acid is derived (e.g., brain, heart, liver,
spleen, etc.). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule,
can be substantially free of other cellular material or culture medium when produced by

recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

[152] A nucleic acid molecule described herein can be isolated using standard molecular biology techniques and the sequence information provided herein. Nucleic acid molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, *et al.*, (eds.), MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

[153] In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NOs: 1-66, or a portion of this nucleotide sequence (*e.g.*, a fragment that can be used as a probe or primer). A nucleic acid molecule that is complementary to the nucleotide sequence shown SEQ ID NOs: 1-66 is one that is sufficiently complementary that it can hydrogen bond with little or no mismatches, thereby forming a stable duplex.

[154] As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, van der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

[155] Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids, a length sufficient to allow for specific hybridization in the case of nucleic acids and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in

respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. Homologs are nucleic acid sequences or amino acid sequences of a particular gene that are derived from different species.

[156] Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid, as described below. Derivatives or analogs of the nucleic acids of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids of the invention, in various embodiments, by at least about 70%, 80%, or 95% identity (with a preferred identity of 80 95%) over a nucleic acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See *e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below.

[157] As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to substantially no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences.

[158] Stringent conditions are known to those skilled in the art and can be found in Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1 6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non limiting example of stringent hybridization conditions are hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C.

Probe Pair Selection

[159] Genotyping of individuals based on the relative copy number of particular

genomic regions can be accomplished using a collection of probe pairs specifically designed as described below.

[160]   First, a list of genomic regions is obtained with the publically available reference genome build number, chromosome, start, and end coordinates. From this list, a custom track file is generated to be used for creation of a custom user track that can be viewed within a browser, for example, the UCSC genome browser, over the Internet.

[161]   After loading the track file into the genome browser, unmasked raw DNA sequences are downloaded using the genome table browser. This is generally downloaded as a multi-FASTA file containing all the regions sequences together, but sometimes it is downloaded as a single sequence FASTA file. The multi-FASTA file containing the region sequences is manipulated to strip extraneous characters from the header of each FASTA entry, leaving only necessary information such as region name, chromosome, start, and end positions.

[162]   Then the multi-FASTA file of region sequences is split into separate FASTA files, one file for each region. Each region file is then passed to a computer program that creates *in silico* restriction fragments of various sizes, depending on the location of restriction enzyme recognition sites in the sequence. The size of the resulting DNA fragments will vary from about 0.1 kb to about 10 kb, with an approximate average size of 0.5 kb.

[163]   When all of the regions have been processed, the resulting fragment sequences are concatenated into one big file again for loading into a database of DNA fragments. A computer program is used to read the file of region coordinates and the file of DNA fragments in order to create a data structure in the database, linking the restriction fragment sequences to the regions. When all of the fragment sequences have been loaded into the database, the Gauntlet program generates probe pairs, each probe consisting of approximately 35-50 nucleotides, specifically designed to work in the nanoString nCounter® system. Many probe pairs are created per restriction fragment during the process, with the majority failing to meet rigorous specifications for the design of matched probe pairs.

[164]   Those probes that pass the physical structure and matching criteria are loaded into the database to be further scored for fitness within the CNV assay. Every passed probe for the project is sent to a local licensed version of the BLAT program to

determine the location(s) with respect to the publically available reference genome (*e.g.* hg18 or hg19) so that we can unambiguously map the location(s) and determine their potential for cross hybridization to other regions of the genome.

[165] Fitness scoring is performed on the Reporter probe of each probe pair (35-50 nucleotides in length) based on: (i) the length of the restriction fragment the probe sequence is within that we know works best in the nCounter® system. This is based on a mathematical model of performance as determined by in-house experimentation. (ii) the location of the probe within the region to be analyzed. The location score is based on the number of probe pairs the customer has requested per region when they submitted them for the assay design. (iii) the results of the BLAT scoring, which gives us a measure of whether we can unambiguously map the Reporter probe within the same reference genome sequence.

[166] The unique algorithm that was developed for fitness scoring is described in detail below. This process allows for an expansion of the number of zones (bins) per region to facilitate selection of the requested number of probe pairs evenly distributed over the entire region using the probes that will work best in the nCounter® system according to our mathematical model of fitness.

[167] The following is the flow of our unique process of probe selection for the custom CNV assay: The user specifies the value (N) of the number of probes requested per region and the maximum value (MAX) of zones to try for an iterative processing of each region. Data is gathered from the database and arranged into data structures amenable to the process. Then, for every region, and for every iteration from N to MAX, a scoring matrix is created the exact length of the region using a natural mathematical function (absolute value of a sine wave) with the cycles specified relative to N, and increasing by the value (S) at the end of each iteration cycle.

[168] For every Reporter probe, a fitness score is calculated using the value of the sine wave, the value of the fragment length model, and an adjusted BLAT score. As the algorithm passes through the entire region length of sequence from end to end, each probe score is tested against the highest score seen so far in the zone. If the score for the probe at that position is higher than the one saved previously, the probe is replaced with the higher scoring probe. When the algorithm has reached the end of each zone, the highest scoring probe is added to the list of probes to be included in the assay.

[169]   When the algorithm has reached the entire length of the region, the program tests if the number of probes requested (N) has been satisfied, and IF NOT, the value of N is changed to the next incremental value of N by adding S, and the process begins all over again. The cycles continue until the maximum value of zones (MAX) has been reached.

EXAMPLES

Example 1: Quantitation of Copy Number Variations.

[170]   Multiple probes for designed for the X (9) and Y (9) chromosomes, as well as for chromosome 18 (6) and chromosome 21(3). Cells lines were provided with known chromosomal aberrations. Samples included cells with 1-5 copies of the X chromosome, normal males and females, and cells with trisomy at chromosome 21 and 18.

[171]   Genomic DNA from the samples was digested with AluI (1hr @ 100ng/μl). 100ng of genomic DNA was used per assay. The genomic DNA was denatured prior to hybridization (5 min @ 95°C, snap cool 2 min). All experiments were performed in triplicate.

[172]   For samples with cells with 1-5 copies of the X chromosome, the samples were normalized to probes on chromosome 18 (see Figure 1). For trisomy detection, normalization was done to the Y chromosome in the male samples (see Figure 2).

[173]   The assay for multiple X chromosomes was repeated. Multi-X cell line genomic DNA purchased from Coriell cell repository. 1 μg of each sample was digested with 10U of AluI enzyme in a 50 μl digest at 37°C for 2 hours. The digested DNA was heat denatured at 95°C for 5 minutes and snap cooled on ice.

[174]   100 ng each digested/denatured sample was hybridized with the codeset CNV2 under standard nCounter® conditions for 16 hours. In addition to probes to other chromosomes, the CNV2 codeset contains 235 probes with 3 probes designed to bind to chromosome X. Standard conditions for the nCounter® assay are: 30 μl hybridization in 5X SSPE/0.1% Tween 20, 25 pM reporter probe, 100 pM capture probe, 100 ng genomic DNA, 65°C, with a heated lid for 16 hours. Raw counts were normalized to the average of 66 probes to invariant genomic regions, to correct for genome-equivalents input into each assay. Normalized counts were averaged for each sample triplicate. Average counts for each probe are then divided by the corresponding count for the reference sample (NA10851, Coriell Cell Repository) to generate a copy number estimate.

51

[175] Copy number estimate is then averaged for the 3 probes to chromosome X. The graph shows the average copy number estimate and standard deviation for 3 probes to chromosome X (see Figure 3). Figures 1-3 show proof of concept that nanoreporters as described above, can be used to ascertain genomic copy number.

Example 2: Precision Counting of Gene Copies in Normal and Cancerous Colon and Breast Tissues.

[176] Hybridizations were performed in triplicate with normal and tumorous colon and breast tissues, with GPCR and Kinase panels (24 total hybridizations) according to the nCounter® probe pair methods described above. As shown in Figure 4, the genes were binned based on expression level, and the average %CV (coefficient of variance, standard deviation and mean) was calculated for all the genes detected in each bin. The total number of measurements (n) is indicated. Even for the 173 measurements of genes expressed at very low levels, the %CV averaged less than 15%. Likewise, small fractional fold changes can be detected according to these methods as shown in Figure 5. This demonstrates the precision of the nCounter® system in the counting of gene copies.

Example 3: Precision Counting of Autosomal Mutations.

[177] 78 different autosomal loci, consisting of 25 HapMapped CNV regions and other regions of interest, together with 23 invariant regions (one per chromosome), were analyzed with a single CodeSet according to the nCounter® method described above. It is easy to identify the loci that vary in copy number.

[178] Samples were, from top to bottom, NA07019, NA12877, NA10487. All the HapMap overlapping regions are 2X in these individuals. NA12877 was male; notice the single X dose in region with index 102.

Example 4: Probe Pair Selection.

[179] *Sequence Preparation:*

[180] A list of genomic regions was obtained. The list could be produced from many sources. For example, the list came, in some instances, from a commercial source and in others from scrutiny of genomic regions known to be involved in a particular disease of interest through a scientific literature search. The list of regions contained specific data concerning the publically available human reference genome that was used to identify the region, including the build identifier, chromosome number, start, and end coordinates.

[181] From the list of regions, a custom track file was generated to be used for creation of a custom user track that can be viewed, for example, within the UCSC genome browser over the Internet. After loading the track file into the genome browser, unmasked raw DNA sequences were downloaded using the genome table browser. This file was generally downloaded as a multi-FASTA file containing all the regions sequences concatenated together, but sometimes it was downloaded as a single sequence FASTA file if only one region is collected.

[182] The multi-FASTA file containing the specific DNA sequences from the regions is manipulated to strip extraneous characters from the header of each FASTA entry, leaving only necessary information such as region name, chromosome, start, and end positions. This can be accomplished using the custom Perl program replace.pl.

[183] The multi-FASTA file of region sequences were then split into separate files, one file for each region using the custom Perl program splitfa.pl. Each file was named using the sequence name found in the header of the FASTA entry in the multi-FASTA file.

[184] Each FASTA sequence file was then passed to the custom Perl program fragment.pl within a Linux bash shell script. The fragment.pl program created *in silico* restriction fragments of various sizes, depending on the location of restriction enzyme recognition sites in the sequence. The size of the resulting DNA fragments varied from about 0.1 kb to about 10 kb, with an approximate average size of 0.5 kb. Parameters set within fragment.pl allowed for the exclusion of fragments smaller or larger than the desired length.

[185] When all of the sequence files had been processed, the resulting sequence fragments were concatenated into one larger file for loading into a database of DNA fragments by the custom Perl program milk.pl. This program connected to a local mySQL CNV database, to create the appropriate project within the database, and linked the data to the specific organism and genome build. It then proceeded to load the DNA sequence fragments into the database for the design of probes. The milk.pl program also read a file of region coordinates in order to create the correct data structure in the database, linking the restriction fragment sequences to the regions.

[186] *Probe Design:*

[187] When all of the fragment sequences had been loaded into the database, the custom Perl program CNV_start_pipeline.pl ran multiple instances of the custom Perl

53

program gauntlet.pl, that generated probe pairs specifically designed to work in the nanoString nCounter® system. Each specific probe consisted of approximately 35-50 nucleotides, with probe pairs being matched for Tm. Many probe pairs were created for each restriction fragment during the process, with the majority failing to meet rigorous specifications for the design of matched probe pairs. Those probes that pass the physical structure and matching criteria were loaded into the CNV database to be scored for fitness within the CNV assay.

[188] *BLAT Scoring:*

[189] BLAT scoring was performed on each probe pair by concatenating the reporter probe sequence and the ghost probe sequence (each approximately 35-50 nucleotides in length). The combined reporter-ghost probe sequence was sent to a local licensed version of the BLAT program to determine the location(s) with respect to the publically available reference genome (*e.g.* hg18 or hg19). This was done to unambiguously map the probes against the reference genome and determine the potential for cross-hybridization to other regions of the genome. The output of the BLAT program for each probe pair was stored in the CNV database for subsequent scoring.

[190] The BLAT score for each probe pair within a project was accomplished by running the custom Perl program CNV_BLAT_score.pl, that created a score in the arbitrary range of 0-100 based on the number of "hits" within a specified combination of "percent identity" (PID) and "alignment length" (AL), parameters set within the program, the location of the "hit" on the combined reporter-ghost probe sequence, and the expected number of "hits" for a particular region in the genome. Note that "hits" within the reporter probe portion of the combined sequence are weighted as more severe than those on the ghost probe portion due to technical considerations of detecting the labeled "spot" tag on the reporter probe.

[191] *Probe Selection:*

[192] When BLAT scores had been determined and loaded into the CNV database, the selection of the "best" probes for each region was accomplished using the custom Perl program butter.pl. This program took into consideration the following parameters for picking the best probes for a given region: (i) the length of the restriction fragment the probe sequence is within that we know works best in the nCounter® system. This was based on a mathematical model of performance as determined by in-house

experimentation, (ii) the location of the probe within the region to be analyzed, and (iii) the BLAT score as determined above in "BLAT scoring".

[193]   The custom Perl program butter.pl used a unique algorithm that was developed for fitness scoring of probe pairs. The algorithm included a process that allows for an expansion of the number of zones (bins) per region to facilitate selection of the requested number of probe pairs evenly distributed over the entire region using the probes that would work best in the nCounter® system according to our mathematical model of fitness. The location score was based on the number of probe pairs that had been requested per region when they submitted them for the assay design. For example, if the desired number of probes to a region was 1, then the best probe pair would be the one that (i) resides on the optimal size restriction fragment (high fragment score), (ii) is directly in the center (equal spacing from either end) of the defined region (high spacing score), and (iii) has the expected number of "hits" (high BLAT score).

[194]   The following is the flow of our unique process of probe selection for the custom CNV assay:

[195]   The user specified the value of the number of probes requested per region (N) and the maximum value of zones to try for an iterative processing of each region (MAX). Data was gathered from the database and arranged into data structures amenable to the process. For each region the following steps were performed.

[196]   For each iteration from N to MAX, a scoring matrix was created the exact length (L) of the region (in nucleotides). Then, using a natural mathematical function (absolute value of a sine wave) with the cycles specified relative to N, a value ranging from 0 to 1 was assigned to every position of the matrix. For example, if the length of the region was 1kb (1000 nucleotides) and the number of desired probes was 1, then L=1000 and N=1, then the value in the matrix position 1000/2 = 500 is equal to 1.0 (the upper limit of the absolute value of a sine function). Therefore, the best position score possible for a probe pair would be at position 500. If a probe pair were located at position 500 with all other parameters being equal amongst the other probe pairs available for the zone (ideal fragment length and ideal BLAT score), the 1 best probe pair selected would be the one located right in the center of the region, which is what we want.

[197]   For each zone (N), the following steps were performed. For each probe pair, a fitness score was calculated using the value of the fragment length model, the value of

the sine wave at the position of the probe pair, and the BLAT score. As the algorithm passed through the entire region length of sequence from end to end, each probe pair score was tested against the highest score seen so far in the zone (there are N zones). If the score for the probe pair at that position was higher than the one saved previously, the probe pair was replaced with the higher scoring one. When the algorithm reached the end of each zone, the highest scoring probe pair is added to the list of those to be included in the assay. Then the next probe pair was scored.

[198]   For the next zone, when the algorithm reached the entire length of the region, the program tests if the number of probes requested (N) has been satisfied, and IF NOT, the value of N is changed to the next incremental value of N by adding S, and the process begins all over again. The cycles continue until the maximum value of zones (MAX) has been reached. In the next iteration (N =N+S unless N=MAX). The value (N) is increased by (S) at the end of every iteration, until MAX is reached. This was then perfomed for each successive region.

Example 5.   Gene Copy Number Assay Protocol.

[199]   *AluI Restriction Digestion*

[200]   A restriction enzyme fragmentation was set up. 200-600 ng of DNA in in 7 μL of sample was placed a 0.2 mL-0.5 mL PCR tube. Genomic DNA was generally at a concentration of at least 29 ng/μL prior to addition to the fragmentation reaction.

[201]   7 μL containing DNA 600 ng was dissolved in RNase free water, Tris pH 8.0 or similar. To this was added 1 μL 10X AluI Fragmentation Buffer; 1 μL 10X CNV DNA Prep Control; 1 μL AluI Fragmentation Enzyme; up to 10 μL of total volume.

[202]   The digestion was incubated at at 37° C for at least 2 hours in a heat block or thermalcycler with heated lid turned on. Samples generally were denatured upon completion of restriction digestion.

[203]   The final hybridization reaction contained the following components: 10 μL nanoreporters, 10 μL hybridization buffer, a total volume of 5 μL of sample DNA, and 5 μL Capture ProbeSet.

[204]   *Hybridization Reaction*

[205]   Aliquots of both reporter probes and capture probes were removed from the freezer and thawed on benchtop at room temperature. The tubes were inverted or flicked several times to mix well and spin down reagent. A master mix was created containing

130 µL of the reporter probes and 130 µL of hybridization buffer. Hybridization buffer was added directly to nanoreporter tube. Capture probes were not added to the master mix. The master mix was inverted to mix and spun down.

[206]   12 tubes were labeled. 20 µL of mastermix was added to each of the 12 tubes.

[207]   *Thermocycler*

[208]   AluI digestion tubes were denatured at 95°C for 5 minutes. The thermocycler was pre-heated to 65°C with heated lid turned on and "forever" time setting. The AluI digestion tubes were then immediately placed on ice for 2 minutes to minimize DNA renaturation and then briefly spun down.

[209]   5 µL AluI-digested DNA sample was addec to the hybridization tube. The remainder of the AluI-digested sample was generally stored at –20° C for future use.

[210]   5 µL of capture probe was added to each tube immediately before placing at 65°C. Tubes were capped and mixed the reagents by flicking with your finger to ensure complete mixing. Tubes were briefly spin down at < 1000 g and immediately placed in the 65°C thermocycler. The hybridization assays were incubated for at least 16 hours and were left at 65°C until ready for processing. Maximum hybridization time did not exceed 30 hours.

Example 6. Optimal Size Selection of Restriction Fragments.

[211]   During the probe design process, target regions were analyzed for AluI sites (based on the human reference genome sequence) and the sizes of AluI fragments were predicted. One nCounter® CNV probe was designed per AluI fragment. Figure 7 shows the relationship between AluI fragment length in base pairs (x-axis) and the counts obtained via hybridization in the nCounter® system (y-axis). Based on this data, the probe design algorithm selected the optimal size fragments per genomic region.

Example 7. Determination of Copy Number Variations (CNV) using nCounter® copy number assay.

[212]   The nCounter® copy number assay was used to measure the copy number of 20 genomic regions in 50 human genomic DNA samples containing known copy number variations (CNVs). Fifty human genomic DNAs were purchased from Coriell Cell Repository. The presence of CNVs in these samples was previously determined by genome-wide analysis using microarray methodology and the data is publically available at the public website Database of Genomic Variants (http://projects.tcag.ca/variation/).

[213] Six hundred nanograms of genomic DNA were processed via the nCounter® Copy number assay manual. Briefly, 600ng of DNA was digested with restriction enzyme Alu1, sample was denatured at 95° C for 5 minutes and added directly to an nCounter® hybridization reaction. The nCounter® CodeSet used in this experiment contained 60 probes complementary to 20 genomic regions that were selected based on known variations across these samples as well as control probes. Hybridization occurred in solution for 16 hours at 65° C. Hybridized samples were purified and imaged using nCounter® PrepStation and Digital Analyzer.

[214] Raw data was normalized using the average signal for 10 invariant control probes. This step removes slight variations in DNA input amounts. Copy number calls were calculated by dividing the normalized counts of the test samples by the normalized counts in the reference and multiplying by 2 for autosomal chromosomes (1-22). The reference sample was NA10851 purchased from Coriell. The graph in Figure 8 shows a comparison of the copy number value (Y-axis) for a region of chromosome 7 between the nCounter® assay (light gray bars) and the public data (dark gray bars, determined by microarray). The nCounter® copy number calls were determined by averaging the copy number calls for the 3 probes designed to this region. There was 100% concordance in copy number calls in this region, across the 50 samples. Several examples of samples with heterozygous (1 copy) and homozygous deletions (zero copies) in this region are shown.

Example 8. Karotype Panel.

[215] The Human Karyotype Panel used below consists of 338 probe pairs designed to target known invariant regions for molecular karyotyping of the human genome. The panel has 8 probe pairs distributed across each arm of the 22 autosomes, excluding the p-arms of acrocentric chromosomes, for a total of 313 autosomal probe pairs. In addition, there are 16 X-specific and 9 Y-specific probe pairs that may be used for gender determination. The Human Karyotype Panel is provided with 10 invariant control probe pairs designed to well-characterized invariant regions on separate chromosomes to be used for normalization of digital count data, eliminating sample-to-sample variation.

[216] Figure 9 shows copy number calls (y axis) for all 313 autosomal probe pairs (x axis) within the Human Karyotype Panel. One hundred and two (102) "normal" human HapMap DNA samples purchased from the Coriell Institute for Medical Research were

tested using the standard CNV assay protocol, described above. Normalized digital counts were compared to the male HapMap sample NA10851 to determine relative copy number calls. For each HapMap sample, the copy number call value is shown as a black cross. Lines marked as "1" represent a z-score of 1 (P<0.32) and lines marked as "2" represent a z-score of 2 (P<0.05) assuming a Gaussian distribution about the expected copy number of 2 and standard deviation of 0.2. The standard deviation of copy number calls for each probe pair is shown at the bottom of the graph, overlaying the y axis for both copy number calls and standard deviation. Probe pairs are sorted by ascending standard deviation. A cut-off value of 0.2 for standard deviation is demarcated by a gray dotted line.

OTHER EMBODIMENTS

[217] While the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims. The patent and scientific literature referred to herein establishes the knowledge that is available to those with skill in the art. All United States patents and published or unpublished United States patent applications cited herein are incorporated by reference. All published foreign patents and patent applications cited herein are hereby incorporated by reference. Genbank and NCBI submissions indicated by accession number cited herein are hereby incorporated by reference. All other published references, documents, manuscripts and scientific literature cited herein are hereby incorporated by reference. While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.

CLAIMS

What is claimed is:

1.   A method of detecting a DNA sequence in a genome comprising:

   (a)   providing a first sample containing genomic DNA;

   (b)   fragmenting the genomic DNA;

   (c)   denaturing the genomic DNA;

   (d)   providing a first nanoreporter comprising a first probe comprising

         (i)   a first label attachment region to which are attached one or more label monomers that emit light constituting a first signal;

         (ii)   a second label attachment region, which is non-over-lapping with the first label attachment region, to which are attached one or more label monomers that emit light constituting a second signal; and

         (iii)   a first target-specific sequence attached to the first probe, wherein the target-specific sequence specifically hybridizes to the genomic DNA sequence to be detected;

         (e) contacting the first probe with the fragmented genomic DNA wherein the contact is made under conditions sufficient for hybridization of the first target specific sequence to a fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected

         (f) stretching the first probe hybridized to the at least one fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected using a flow-stretch, receding meniscus, or electro-stretch technique, thereby spatially separating said label monomers, and

         (g) measuring a signal from the first probe, wherein said signal uniquely identifies the at least one fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected.

2.   The method of claim 1, wherein the first nanoreporter further comprises a second probe comprising

         (i)   a second target-specific sequence; and

(ii)   an affinity tag;

wherein the first probe and the second probe specifically hybridize to the same fragment
of the fragmented genomic DNA at different sites on the fragment.

3.   The method of claim 1, wherein the first probe further comprises an affinity tag.

4.   The method of claim 1, wherein the genomic DNA is mammalian genomic DNA.

5.   The method of claim 4, wherein the mammal is a human.

6.   The method of claim 1, wherein the genomic DNA sample is unamplified.

7.   The method of claim 1, wherein step (e) is performed in solution.

8.   The method of claim 1, wherein the fragmentation is performed by restriction
enzyme digestion.

9.   The method of claim 8, wherein the restriction enzyme is AluI.

10.   The method of claim 8, wherein the restriction enzyme is BfaI.

11.   The method of claim 1, wherein the fragmentation is performed chemically, by
mechanical shearing or sonication.

12.   A method of determining the copy number of the DNA sequence to be detected
of claim 1 further comprising

(a) providing a reference sample comprising fragmented genomic DNA wherein
the copy number of the genomic sequence the first target specific sequence specifically
hybridizes to in the reference sample is known;

(b) contacting the first probe with the reference sample wherein the contact is
made under conditions sufficient for hybridization of the first target specific sequence to
a fragment of the fragmented genomic DNA comprising the genomic DNA sequence to

be detected;

(c) stretching the first probe hybridized to the at least one fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected using a flow-stretch, receding meniscus, or electro-stretch technique, thereby spatially separating said label monomers;

(d) measuring a signal from the first probe, wherein said signal uniquely identifies the at least one fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected; and

(e) comparing the signal from the first sample to the signal from the reference sample, wherein the copy number of the first sample is determined by correlating the signal from the first sample with the signal from the reference sample.

13.   The method of claim 12, wherein the reference sample is a synthetic nucleic acid sample.

14.   The method of claim 12, wherein the reference sample is a biological genomic DNA sample.

15.   A method of normalizing the signal generated in claim 12 further comprising

(a)   providing at least a second nanoreporter comprising a third probe comprising

(i)   a third label attachment region to which are attached one or more label monomers that emit light constituting a third signal;

(ii)   a fourth label attachment region, which is non-over-lapping with the third label attachment region, to which are attached one or more label monomers that emit light constituting a fourth signal; and

(iii)   a third target-specific sequence attached to the third probe, wherein the target-specific sequence specifically hybridizes to a first DNA fragment from a copy number invariant region of the genome;

(b) contacting the third probe with the fragmented genomic DNA from the first sample and the reference sample wherein the contact is made under conditions sufficient for hybridization of the third target specific sequence to the first DNA fragment from a

copy number invariant region of the genome;

(c) stretching the third probe hybridized to the first DNA fragment from a copy number invariant region of the genome using a flow-stretch, receding meniscus, or electro-stretch technique, thereby spatially separating said label monomers;

(d) measuring a signal from the third probe, wherein said signal uniquely identifies the first DNA fragment from a copy number invariant region of the genome; and

(e) comparing the signal from the the second nanoreporter contacted with the first sample and the second nanoreporter contacted with the reference sample, wherein the number of multiples of the quantity of signal from the second nanoreporter contacted with the first sample compared to the quantity of signal from the second nanoreporter contacted with the reference sample normalizes the signal from the first nanoreporter contacted with the first sample.

16.   The method of claim 15, wherein the second nanoreporter further comprises a fourth probe comprising

(i)   a fourth target-specific sequence; and

(ii)   an affinity tag;

wherein the third probe and the fourth probe specifically hybridize to the same first DNA fragment from a copy number invariant region of the genome; at different sites on the fragment.

17.   The method of claim 15, wherein the first DNA fragment from a copy number invariant region of the genome comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1-66.

18.   The method of claim 15, wherein the first DNA fragment from a copy number invariant region of the genome comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 2, 5, 7, 12, 13, 17, 19, 24, 25, 28, 32, 36, 38, 40, 44, 46, 50, 52, 56, 58, 62 and 66.

19.   The method of claim 15, wherein the first DNA fragment from a copy number

invariant region of the genome e comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 2, 5, 13, 19, 28, 46, 50, 56, 58 and 66.

20.   The method of claim 15, wherein the third probe further comprises an affinity tag.

21.   The method of claim 15, wherein the genomic DNA is mammalian genomic DNA.

22.   The method of claim 21, wherein the mammal is a human.

23.   The method of claim 1, wherein the signal generated from the first nanoreporter hybridized to the at least one fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected comprises a mixture of two or more different label monomers.

24.   The method of claim 15, wherein the signal generated from the second nanoreporter hybridized to the first DNA fragment from a copy number invariant region of the genome comprises a mixture of two or more different label monomers.

25.   The method of claim 1, wherein said labels are fluorescent.

26.   The method of claim 15, wherein said labels are fluorescent.

27.   A method of detecting a DNA sequence in a genome comprising:
    (a)   providing a first sample containing genomic DNA;
    (b)   fragmenting the genomic DNA;
    (c)   denaturing the genomic DNA;
    (d)   providing a first nanoreporter that specifically hybridizes to the genomic DNA sequence to be detected;
    (e) contacting the first nanoreporter with the fragmented genomic DNA wherein the contact is made under conditions sufficient for hybridization of the first nanoreporter

to the genomic DNA sequence to be detected;

(f) stretching the first nanoreporter hybridized to the at least one fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected using a flow-stretch, receding meniscus, or electro-stretch technique; and

(g) measuring a signal from the first nanoreporter, wherein said signal uniquely identifies the at least one fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected.

28.   A method of determining the copy number of the DNA sequence to be detected of claim 27 further comprising

(a) providing a reference sample comprising fragmented genomic DNA wherein the copy number of the genomic sequence the first nanoreporter specifically hybridizes to in the reference sample is known;

(b) contacting the first nanoreporter with the reference sample wherein the contact is made under conditions sufficient for hybridization of the first nanoreporter to a fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected;

(c) stretching the first nanoreporter hybridized to the at least one fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected using a flow-stretch, receding meniscus, or electro-stretch technique,

(d) measuring a signal from the first nanoreporter, wherein said signal uniquely identifies the at least one fragment of the fragmented genomic DNA comprising the genomic DNA sequence to be detected; and

(e) comparing the signal from the first sample to the signal from the reference sample, wherein the copy number of the first sample is determined by correlating the signal from the first sample with the signal from the reference sample.

29.   A method of normalizing the signal generated in claim 28 further comprising

(a) providing at least a second nanoreporter;

(b) contacting the second nanoreporter with the the fragmented genomic DNA from the first sample and the reference sample wherein the contact is made under conditions sufficient for hybridization of the second nanoreporter toa first DNA fragment

from a copy number invariant region of the genome;

(c) stretching the second nanoreporter hybridized to the first DNA fragment from a copy number invariant region of the genome using a flow-stretch, receding meniscus, or electro-stretch technique;

(d) measuring a signal from the second nanoreporter, wherein said signal uniquely identifies the first DNA fragment from a copy number invariant region of the genome; and

(e) comparing the signal from the the second nanoreporters contacted with the first sample and the second nanoreporter contacted with the reference sample, wherein the number of multiples of the quantity of signal from the second nanoreporter contacted with the first sample compared to the quantity of signal from the second nanoreporter contacted with the reference sample normalizes the signal from the first nanoreporter contacted with the first sample.

30.   The method of claim 29, wherein the first DNA fragment from a copy number invariant region of the genome comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1-66.

31.   A method of detecting two or more DNA sequences in a genome comprising:

(a)   providing a first sample containing genomic DNA;

(b)   fragmenting the genomic DNA;

(c)   denaturing the genomic DNA;

(d)   providing a two or more nanoreporters that each specifically hybridize to a distinct genomic DNA sequence to be detected;

(e) contacting the two or more nanoreporters with the fragmented genomic DNA wherein the contact is made under conditions sufficient for hybridization of the two or more nanoreporters to the genomic DNA sequence to be detected;

(f) stretching the two or more nanoreporters hybridized to its corresponding distinct genomic DNA sequence using a flow-stretch, receding meniscus, or electro-stretch technique, and

(g) measuring a signal from the two or more nanoreporters, wherein said signal uniquely identifies each of the corresponding distinct genomic DNA sequences thereby

detecting two or more DNA sequences in a genome.

32.   A method of determining the copy number of the DNA sequence to be detected of claim 31 further comprising

(a) providing a reference sample comprising fragmented genomic DNA wherein the copy number of the genomic sequence the first target specific sequence specifically hybridizes to in the reference sample is known;

(b) contacting the first nanoreporter with the reference sample wherein the contact is made under conditions sufficient for hybridization of the two or more nanoreporters each fragment of the fragmented genomic DNA comprising the genomic DNA sequences to be detected;

(c) stretching the two or more nanoreporters hybridized to each fragment of the fragmented genomic DNA comprising the genomic DNA sequences to be detected using a flow-stretch, receding meniscus, or electro-stretch technique,

(d) measuring signals from the two or more nanoreporters, wherein said signal uniquely identifies each of the fragments of the fragmented genomic DNA comprising the genomic DNA sequences to be detected;

(e) comparing the signal from the first sample to the signal from the reference sample, wherein the copy number of the first sample is determined by correlating the signal from the first sample with the signal from the reference sample.

33.   A method of normalizing the signal generated in claim 32 further comprising

(a) providing at least one invariable sequence specific nanoreporter;

(b) contacting the at least one copy number invariant sequence specific nanoreporter with the fragmented genomic DNA from the first sample and the reference sample wherein the contact is made under conditions sufficient for hybridization of the at least one invariable sequence specific nanoreporter to a first DNA fragment from a copy number invariant region of the genome;

(c) stretching the second nanoreporter hybridized to the the first DNA fragment from a copy number invariant region of the genome using a flow-stretch, receding meniscus, or electro-stretch technique;

(d) measuring a signal from the second nanoreporter, wherein said signal

uniquely identifies the first DNA fragment from a copy number invariant region of the genome; and

(e) comparing the signal from the the second nanoreporters contacted with the first sample and the second nanoreporter contacted with the reference sample, wherein the number of multiples of the quantity of signal from the second nanoreporter contacted with the first sample compared to the quantity of signal from the second nanoreporter contacted with the reference sample normalizes the signal from the first nanoreporter contacted with the first sample.

34. The method of claim 33, wherein the first DNA fragment from a copy number invariant region of the genome comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1-66.

35. A kit, comprising

(a) a first nanoreporter comprising a first probe comprising

(i) a first label attachment region to which are attached one or more label monomers that emit light constituting a first signal;

(ii) a second label attachment region, which is non-over-lapping with the first label attachment region, to which are attached one or more label monomers that emit light constituting a second signal; and

(iii) a first target-specific sequence attached to the first probe, wherein the target-specific sequence specifically hybridizes to a target DNA sequence; and

(b) a restriction enzyme.

36. A composition comprising an isolated nucleic acid molecule comprising at least 50 nucleotides of a sequence selected from the group consisting of SEQ ID NO: 1-66.

37. A method of selecting probe pairs for detection of a genomic sequence comprising:

(a) providing the genomic sequence;

(b) performing *in silico* restriction fragmentation of the genomic sequence;

(c) generating *in silico* probe pairs for every position on the *in silico*

restriction fragments, wherein each member of each pair is 35-50 nucleotides in length, and wherein each member of each set of pairs is complementary to a contiguous sequence;

     (d)   discarding sets of probe pairs wherein the melting temperatures of the probe pairs differ by more than 5° C;

     (e)   subjecting the remaining probe pairs to BLAT scoring; and

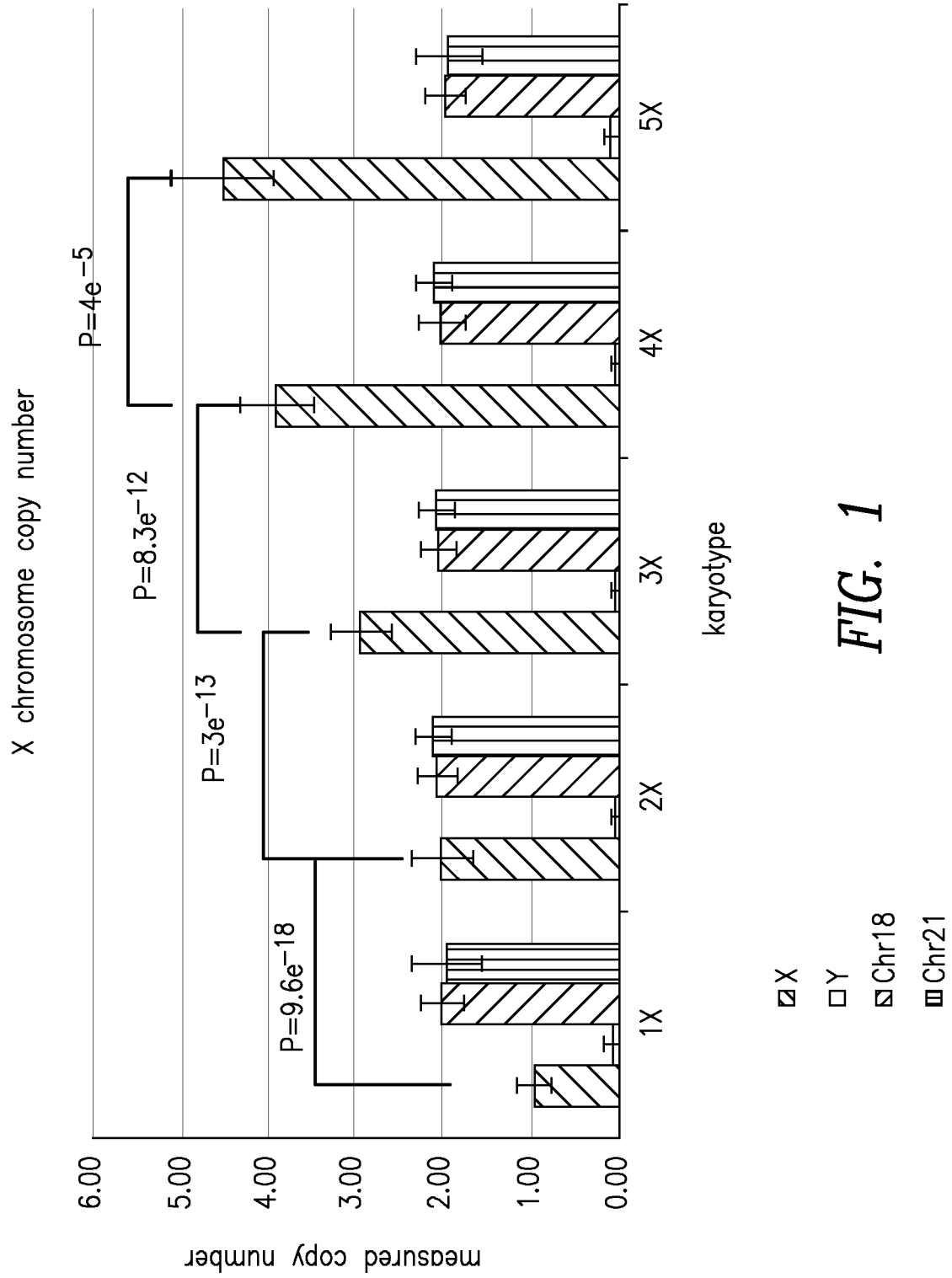   (f)   discarding sets of probe pairs with the lowest 75% of BLAT scores; thereby selecting probe pairs for detection of a genomic sequence.

FIG. 1

FIG. 2

FIG. 3

FIG. 4

FIG. 5

FIG. 6 1-1

7/14

FIG. 6 1-2

FIG. 6 2-1

FIG. 6-2-2

10/14

FIG. 6-3-1

11/14

FIG. 6-3-2

FIG. 7

FIG. 8

14/14

FIG. 9