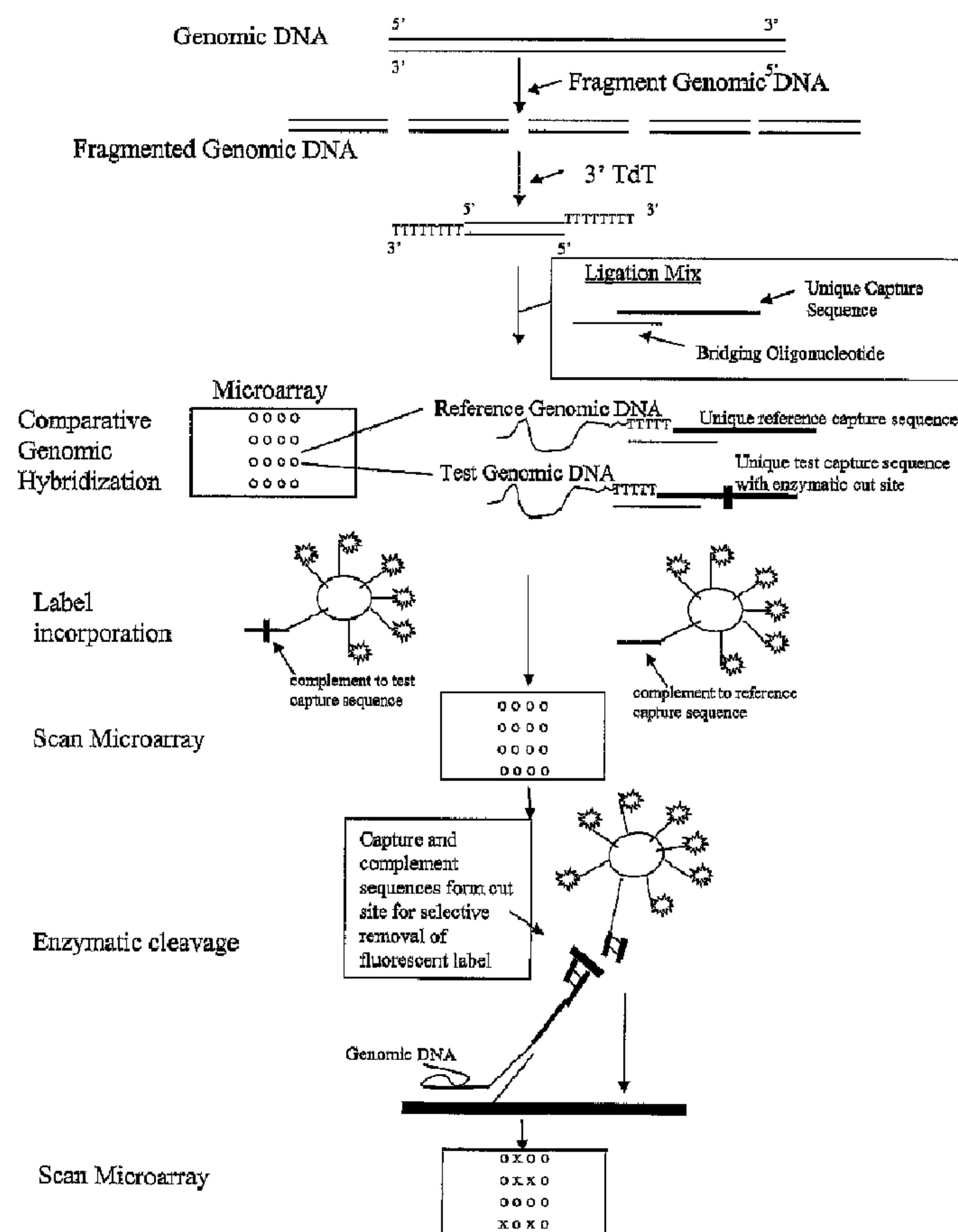




(86) Date de dépôt PCT/PCT Filing Date: 2005/05/19
(87) Date publication PCT/PCT Publication Date: 2005/12/01
(45) Date de délivrance/Issue Date: 2014/09/30
(85) Entrée phase nationale/National Entry: 2006/11/20
(86) N° demande PCT/PCT Application No.: US 2005/017986
(87) N° publication PCT/PCT Publication No.: 2005/113834
(30) Priorités/Priorities: 2004/05/20 (US60/573,457);
2005/01/27 (US60/648,374)

(51) Cl.Int./Int.Cl. *C12Q 1/68* (2006.01)
(72) Inventeurs/Inventors:
MOHAMMED, MANSOOR S., US;
DZIDIC, NATASA, US;
MCCASKILL, CHRISTOPHER, US;
KIM, JAEWEON, US
(73) Propriétaire/Owner:
QUEST DIAGNOSTICS INVESTMENTS
INCORPORATED, US
(74) Agent: GOWLING LAFLEUR HENDERSON LLP

(54) Titre : HYBRIDATION COMPARATIVE A MARQUEUR UNIQUE
(54) Title: SINGLE LABEL COMPARATIVE HYBRIDIZATION



(57) Abrégé/Abstract:

The present invention provides methods of detecting and mapping chromosomal or genetic abnormalities associated with various diseases or with predisposition to various diseases, or to detecting the phenomena of large scale copy number variants. In

(57) Abrégé(suite)/Abstract(continued):

particular, the present invention provides advanced methods of performing array-based comparative hybridization that allow reproducibility between samples and enhanced sensitivity by using the same detectable label for both test sample and reference sample nucleic acids. Invention methods are useful for the detection or diagnosis of particular disease conditions such as cancer, and detecting predisposition to cancer based on detection of chromosomal or genetic abnormalities and gene expression level. Invention methods are also useful for the detection or diagnosis of hereditary genetic disorders or predisposition thereto, especially in prenatal samples. Moreover, invention methods are also useful for the detection or diagnosis of de novo genetic aberrations associated with post-natal developmental abnormalities.

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

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
1 December 2005 (01.12.2005)

PCT

(10) International Publication Number
WO 2005/113834 A2

(51) International Patent Classification⁷: **C12Q 1/68**

(21) International Application Number:
PCT/US2005/017986

(22) International Filing Date: 19 May 2005 (19.05.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/573,457 20 May 2004 (20.05.2004) US
60/648,374 27 January 2005 (27.01.2005) US

(71) Applicant (for all designated States except US): **QUEST DIAGNOSTICS INVESTMENTS INCORPORATED** [US/US]; 300 Delaware Avenue, Wilmington, DE 19899 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **MOHAMMED, Mansoor, S.** [—/US]; Mission Viejo. **DZIDIC, Natasa**

[—/US]; Laguna Beach, CA. **MCCASKILL, Christopher** [US/US]; 2000 Corporate Drive #906, Ladera Ranch, CA 92694 (US). **KIM, Jaeweon** [KR/US]; 976 Sea Cliff Drive, Carlsbad, CA 92009 (US).

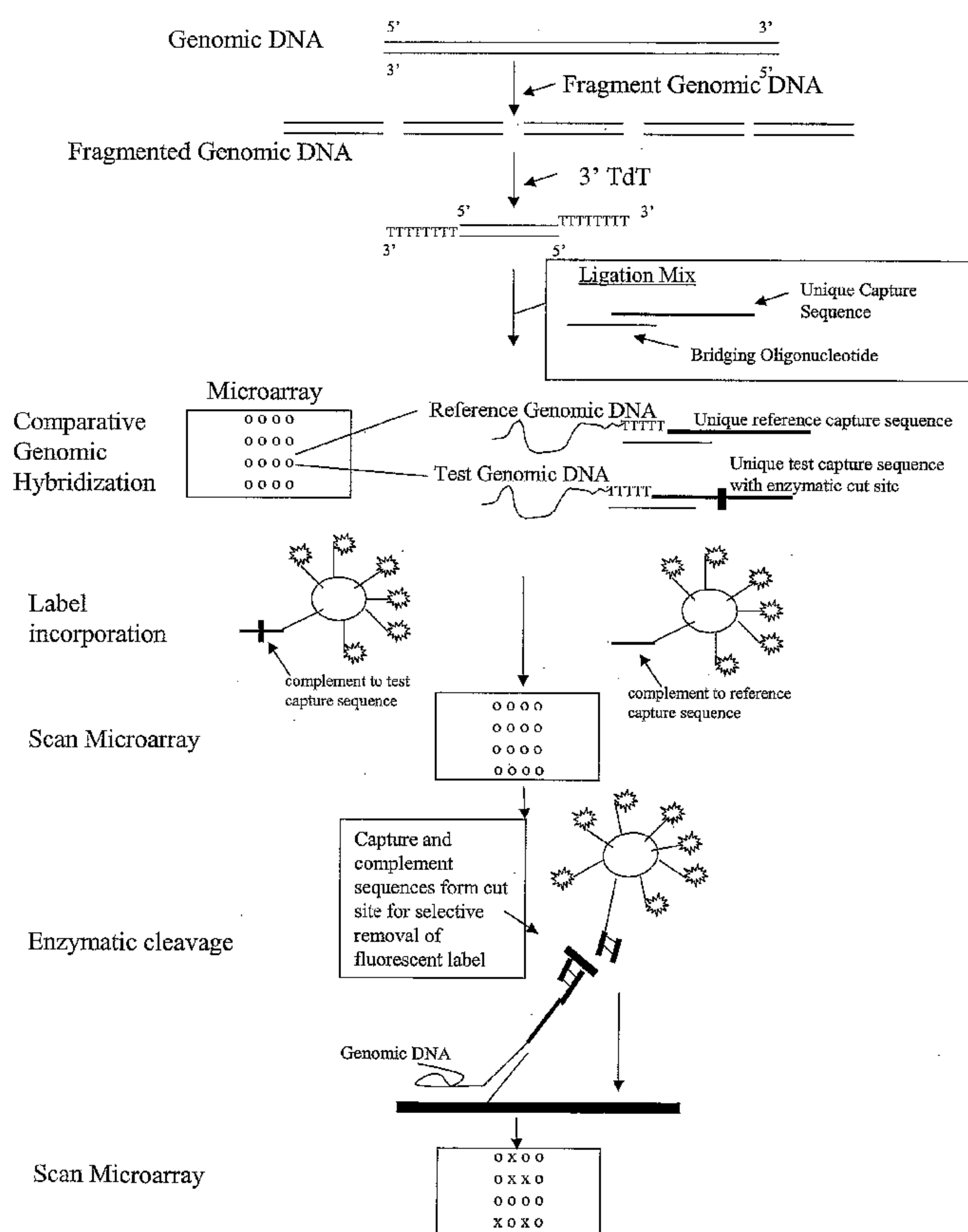
(74) Agents: **WARBURG, Richard, J.** et al.; Foley & Lardner L.L.P., P.O. Box 80278, San Diego, CA 92138-0278 (US).

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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH,

[Continued on next page]

(54) Title: SINGLE LABEL COMPARATIVE HYBRIDIZATION



(57) Abstract: The present invention provides methods of detecting and mapping chromosomal or genetic abnormalities associated with various diseases or with predisposition to various diseases, or to detecting the phenomena of large scale copy number variants. In particular, the present invention provides advanced methods of performing array-based comparative hybridization that allow reproducibility between samples and enhanced sensitivity by using the same detectable label for both test sample and reference sample nucleic acids. Invention methods are useful for the detection or diagnosis of particular disease conditions such as cancer, and detecting predisposition to cancer based on detection of chromosomal or genetic abnormalities and gene expression level. Invention methods are also useful for the detection or diagnosis of hereditary genetic disorders or predisposition thereto, especially in prenatal samples. Moreover, invention methods are also useful for the detection or diagnosis of de novo genetic aberrations associated with post-natal developmental abnormalities.

WO 2005/113834 A2



GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

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

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

SINGLE LABEL COMPARATIVE HYBRIDIZATION

FIELD OF THE INVENTION

[0001] The present invention relates to the detection and mapping of chromosomal or genetic abnormalities associated with various diseases or with predisposition to various diseases. In a particular aspect, the present invention relates to the use of nucleic acids in comparative hybridization.

BACKGROUND OF THE INVENTION

[0002] The following description is provided to assist the understanding of the reader. None of the information provided or references cited is admitted to be prior art to the present invention.

[0003] Comparative hybridization methods test the ability of two nucleic acids to interact with a third target nucleic acid. In particular, comparative genomic hybridization (CGH) is a method for detecting chromosomal abnormalities. CGH was originally developed to detect and identify the location of gain or loss of DNA sequences, such as deletions, duplications or amplifications commonly seen in tumors (Kallioniemi *et al.*, Science 258:818-821, 1992). For example, genetic changes resulting in an abnormal number of one or more chromosomes (*i.e.*, aneuploidy) have provided useful diagnostic indicators of human disease, specifically as cancer markers. Changes in chromosomal copy number are found in nearly all major human tumor types. For a review, see Mittelman *et al.*, "Catalog of Chromosome Aberrations" in CANCER, Vol. 2 (Wiley-Liss, 1994).

[0004] In addition, the presence of aneuploid cells has also been used as a marker for genetic chromosomal abnormalities. Various chromosomal abnormalities may occur in an estimated 0.5% of all live births. For example, Down's syndrome or trisomy 18 which has an incidence of about 1 in 800 live births, is commonly the subject of a variety of prenatal screens or diagnostic techniques. Chromosomal aneuploidies involving chromosomes 13, 18, 21, X and Y account for up to 95% of all liveborn chromosomal aberrations resulting in birth defects (Whiteman *et al.*, *Am. J. Hum. Genet.* 49:A127-129, 1991), and up to 67% of all chromosomal abnormalities, including balanced translocations (Klinger *et al.*, *Am. J. Hum. Genet.* 51:52-65, 1992).

[0005] CGH is useful to discover and map the location of genomic sequences with variant copy number without prior knowledge of the sequences. Oligonucleotide probes directed to known mutations are not required for CGH. Early CGH techniques employ a competitive *in situ* hybridization between test DNA and normal reference DNA, each labeled with a different color, and a metaphase chromosomal spread. Chromosomal regions in the test DNA, which are at increased or decreased copy number as compared to the normal reference DNA can be quickly identified by detecting regions where the ratio of signal from the two different colors is altered. For example, those genomic regions that have been decreased in copy number in the test cells will show relatively lower signal from the test DNA than the reference (compared to other regions of the genome (*e.g.*, a deletion)); while regions that have been increased in copy number in the test cells will show relatively higher signal from the test DNA (*e.g.*, a duplication). Where a decrease or an increase in copy number is limited to the loss or gain of one copy of a sequence, CGH resolution is usually about 5-10 Megabases (Mb).

[0006] CGH has more recently been adapted to analyze individual genomic nucleic acid sequences rather than a metaphase chromosomal spread. Individual nucleic acid sequences are arrayed on a solid support, and the sequences can represent the entirety of one or more chromosomes, or the entire genome. The hybridization of the labeled nucleic acids to the array targets is detected using different labels, *e.g.*, two color fluorescence. Thus, array-based CGH with a plurality of individual nucleic acid sequences allows one to gain more specific information than a chromosomal spread, is potentially more sensitive, and facilitates the analysis of samples.

[0007] For example, in a typical array-based CGH, equitable amounts of total genomic nucleic acid from cells of a test sample and a normal reference sample are labeled with two different colors of fluorescent dye and co-hybridized to an array of BACs, which contain the cloned nucleic acid fragments that collectively cover the cell's genome. The resulting co-hybridization produces a fluorescently labeled array, the coloration of which reflects the competitive hybridization of sequences in the test and reference genomic DNAs to the homologous sequences within the arrayed BACs. Theoretically, the copy number ratio of homologous sequences in the test and reference genomic nucleic acid samples should be directly proportional to the ratio of their respective colored fluorescent signal intensities at discrete BACs within the array. Array-based CGH is described in U.S. Patent Nos. 5,830,645

and 6,562,565 for example, using target nucleic acids immobilized on a solid support in lieu of a metaphase chromosomal spread.

[0008] When combining more than one color or type of labeled nucleic acid in a hybridization mixture, the relative concentrations and/or labeling densities may be adjusted for various purposes. Adjustments may be made by selecting appropriate detection reagents (avidin, antibodies and the like), or by the design of the microscope filters among other parameters. When using quantitative image analysis, mathematical normalization can be used to compensate for general differences in the staining intensities of different colors. Thus, the use of different labels to distinguish test from reference genomic nucleic acids in traditional CGH entails additional refinements or adjustments that complicate sample processing, standardization across samples, and evaluation of the results obtained. For example, when using visual observation or photography of the results, the individual color intensities need to be adjusted for optimum observability of changes in their relative intensities.

[0009] As described below, improvements in comparative hybridization methods including CGH are provided.

SUMMARY OF THE INVENTION

[0010] It is an object of the present invention to overcome the limitations of using different labels in comparative hybridization (for example, different fluorescent colors that must be read at two different wavelengths) and in particular, comparative genomic hybridization methods. Accordingly, provided is a method of performing comparative hybridization by comparing the amount of test and reference nucleic acids hybridized to a nucleic acid array, the amounts determined by detecting a signal from the hybridized nucleic acids which are labeled with the same detectable label. This method is applicable to comparative hybridization methods in general and to CGH in particular. Accordingly, reference to CGH where the test and reference nucleic acid is genomic nucleic acid should be understood to encompass methods where the test and reference nucleic acids are other than genomic nucleic acids. By the same token, it will be understood that the type of label used is not critical and that various labels described herein and known in the art and yet to be discovered may be used in this invention and that reference to a single type label (e.g. fluorescent label) in any of the embodiments of the invention disclosed herein should be understood to include such other types of labels.

[0011] In a preferred embodiment, CGH is performed using two samples of genomic nucleic acids; a test sample containing genomic nucleic acids, and a reference or control sample containing genomic nucleic acids with no known chromosomal or genetic abnormalities. The test sample and the reference sample are co-hybridized to a nucleic acid array that contains a plurality of nucleic acids or nucleic acid segments spotted onto a surface (such as a glass slide) at discrete locations. The array may contain target nucleic acid markers for certain known genetic mutations or disease states, or may represent (in aggregate) an entire chromosome, or the full chromosomal complement to obtain a genetic profile similar to karyotyping. In these approaches the detectable label may be attached to the test and reference nucleic acids before hybridization or after hybridization. In another approach, the detectable label may be attached to one of the test or reference nucleic acids before hybridization while the label is attached to the other of the test or reference nucleic acid after hybridization. The detectable label may be attached covalently or non-covalently such as by a ligand-receptor interaction or by hybridization between complementary nucleotide sequences.

[0012] Central to the invention is the ability to use the same label to determine the relative amount of test and reference nucleic acids hybridized to the array. The present invention achieves this end by various approaches. In what may be referred to as a “subtractive” approach, the linkage used to attach the detectable label to the test nucleic acid and the reference nucleic acid is different, allowing for selective cleavage or removal of one linkage over that of the other. As a first step, the total detectable signal on the array, which represents label linked to both the test sample and the reference sample nucleic acids hybridized to the array, is first positionally quantified. The array is then subjected to a condition or treatment that causes selective cleavage of the linker such that the label is stripped from either the hybridized test or reference nucleic acids, whichever has the susceptible linkage. The remaining signal representing nucleic acid that is not linked to the susceptible linker is then positionally quantified. The next step comprises using the results of the two reads to determine the amount of the hybridized nucleic acid that is attached to the label via the susceptible linkage. In a preferred approach, the signal representing the nucleic acid that is linked to the label by the susceptible linker is determined by subtracting the remaining signal following selective removal from the total signal. The signal from the two samples thus determined can be used to identify differences between the test sample genomic

nucleic acids and the reference sample genomic nucleic acids so as to detect chromosomal or genetic abnormalities associated with the test sample nucleic acid.

[0013] The present invention also achieves comparative hybridization using the same detectable label by another approach that may be referred to as an “additive” approach. In accordance with this approach, the test sample nucleic acids comprise a first tag; and the reference sample nucleic acids comprise a second tag. Following hybridization, the surface is contacted with a first complex containing a detectable label and a first entity, such that the first complex selectively binds with the first tag. The next step comprises determining the location and amount of the detectable label bound to the array surface (*i.e.*, to “read” the array). Once the array is read to determine the amount of detectable label associated with nucleic acid that comprises the first tag, the surface is then contacted with a second complex containing the same detectable label as present in the first complex and containing a second entity, such that the second complex selectively binds with the second tag. The array is then read a second time to determine the location and amount of the total detectable label representing both nucleic acids hybridized to the surface. The last step comprises using the results of the two reads to determine the amount of the hybridized nucleic acid that is associated with the second tag. In a preferred approach, the first read is subtracted from the second read to obtain the signal representing the nucleic acid that is linked to the second tag. The signal from the two samples thus determined can be used to identify differences between the test sample genomic nucleic acids and the reference sample genomic nucleic acids so as to detect any chromosomal or genetic abnormalities associated with the test sample nucleic acid.

[0014] In a variant of this approach, the amount of the hybridized nucleic acid that is associated with the second tag can be determined using a duplicate of the hybridized array but which has not been contacted with the first complex. Thus, the duplicate array is contacted with the second complex and not the first complex. The signal from this second array directly represents the amount of hybridized nucleic acid with the second tag, which can be compared to the amount of signal from the first array that was contacted only with the first complex and represents the amount of hybridized nucleic acid associated with the first tag. Because the two analyses are independent of each other, each array may be processed in any order or simultaneously.

In another approach, one may first hybridize the array with the test and reference nucleic acids wherein one of the test and reference nucleic acids has already been labeled (e.g. by random priming). The array is then read after hybridization to determine signal corresponding with the particular labeled nucleic acid sample. The array is then contacted with a complex comprising a detectable label and an entity, wherein the complex selectively reacts with the other of the test or reference nucleic acid via a tag attached to said other of the test or reference nucleic acid. The assay is read again to measure the total signal for both hybridized nucleic acids. The next step comprises using the results of the two reads to determine the amount of the hybridized nucleic acid that is associated with the tag. In a preferred approach, the first read is subtracted from the second read to obtain the signal representing the nucleic acid that was linked to the tag. The signal from the two samples thus determined can be used to identify differences between the test sample genomic nucleic acids and the reference sample genomic nucleic acids so as to detect any chromosomal or genetic abnormalities associated with the test sample nucleic acid.

[0015] In a variant of this approach, the amount of the hybridized nucleic acid that is associated with the second tag can be determined using a duplicate of the hybridized array except that the duplicate is prepared by hybridizing to test and reference nucleic acids that do not contain a detectable label. In this case, the duplicate array is contacted with a complex comprising a detectable label and an entity, wherein the complex selectively reacts with the other of the test or reference nucleic acid via a tag attached to said other of the test or reference nucleic acid. The signal from this second array directly represents the amount of hybridized nucleic acid with the second tag, which can be compared to the amount of signal from the first array that was contacted only with the first complex and represents the amount of hybridized nucleic acid associated with the first tag. Because the two analyses are independent of each other, each array may be processed in any order or simultaneously.

[0016] The term "tag" as used herein, refers to any physical molecule directly or indirectly associated with a nucleic acids of a sample such that substantially all individual nucleic acid segments of that sample can be marked, purified, or captured via the same tag. The tag may be a member of a specific binding pair such as a ligand-receptor or a pair of oligonucleotides with a complementary sequence. The tag/entity interaction referred to herein is understood to be a specific binding pair such as a ligand/receptor binding pair or a pair oligonucleotides with a complementary sequence. A tag/entity combination should be chosen so that the it does not appreciably interact with the other tag/entity combination that is

used together in comparative hybridization. This allows one to identify hybridized test from the reference nucleic acid by the specific interaction associated with each tag/entity.

[0017] In a preferred embodiment, the tag comprises a unique oligonucleotide “capture sequence,” which refers to a sequence of nucleotides that is essentially unique to the assay. In this case, the reactive entity comprises an oligonucleotide complementary to the unique oligonucleotide capture sequence associated with one sample of nucleic acid segments. Preferably, the reactive entity complementary oligonucleotide is in a dendrimeric construct to provide a multiplicity of the detectable label.

[0018] As used herein, specific binding pair members include antigen-antibody, biotin-avidin, hormone-receptor, receptor-ligand, enzyme-substrate, IgG-protein A, and the like.

[0019] As used herein, “nucleic acid” refers to segments or portions of DNA, cDNA, and/or RNA. Nucleic acid may be derived or obtained from an originally isolated nucleic acid sample from any source (*e.g.*, isolated from, purified from, amplified from, cloned from, reverse transcribed from sample DNA or RNA) or may be synthesized *de novo*. Nucleic acid includes an oligonucleotide, nucleotide or polynucleotide, and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin that may be single or double stranded, and represent the sense or antisense strand.

[0020] “Genomic nucleic acid” refers to some or all of the DNA from the nucleus of a cell. In some embodiments, genomic DNA may include sequence from all or a portion of a single gene or from multiple genes, sequence from one or more chromosomes, or sequence from all chromosomes of a cell. In contrast, the term “total genomic nucleic acid” is used herein to refer to the full complement of DNA contained in the genome of a cell. As is well known, genomic nucleic acid includes gene coding regions, introns, 5' and 3' untranslated regions, 5' and 3' flanking DNA and structural segments such as telomeric and centromeric DNA, replication origins, and intergenic DNA. Genomic nucleic acid may be obtained from the nucleus of a cell, or recombinantly produced. Genomic DNA also may be transcribed from DNA or RNA isolated directly from a cell nucleus. PCR amplification also may be used. Methods of purifying DNA and/or RNA from a variety of samples are well-known in the art.

[0021] As used herein, “cDNA” refers to DNA which is copied from RNA. cDNA copied from mRNA does not include the various non-coding sequences characteristic of genomic DNA.

[0022] As used herein, “chromosomal abnormality” refers to any difference in the DNA sequence from a wild-type or normal cell. A chromosomal abnormality may reflect a difference between the full genetic complement of all chromosomes contained in an organism, or any portion thereof, as compared to a normal full genetic complement of all chromosome in that organism. For example, a chromosomal abnormality may include a change in chromosomal copy number (*e.g.*, aneuploidy), or a portion thereof (*e.g.*, deletions, amplifications); or a change in chromosomal structure (*e.g.*, translocations, mutations). “Aneuploid cell” or “aneuploidy” as used herein, refers to a cell having an abnormal number of at least one chromosome in interphase. For example, a normal human cell contains a total of 46 chromosomes in interphase, or 2 copies of each of chromosomes 1 through 22, and 2 sex chromosomes (XX or XY). An abnormal chromosomal copy number is any number other than two of the normal chromosomal complement of two copies of chromosomes 1 through 22, and any combination other than two of the sex chromosomes X and Y.

[0023] As used herein, “genetic abnormality” refers to a chromosomal abnormality that is known to be associated with a particular disease condition (*e.g.*, a specific gene mutation causing a dysfunctional protein directly causing a disease state). A chromosomal or genetic abnormality may be hereditary, *i.e.*, passed from generation to generation.

[0024] A “sample” as used herein may be acquired from essentially any diseased or healthy organism, including humans, animals and plants, as well as cell cultures, recombinant cells, cell components and environmental sources. Samples may be from any animal, including by way of example and not limitation, humans, dogs, cats, sheep, cattle, and pigs. Samples can be a biological tissue, fluid or specimen. Samples may include, but are not limited to, amniotic fluid, blood, blood cells, cerebrospinal fluid, fine needle biopsy samples, peritoneal fluid, plasma, pleural fluid, saliva, semen, serum, sputum, tissue or tissue homogenates, tissue culture media, urine, and the like. Samples may also be processed, such as sectioning of tissues, fractionation, purification, or cellular organelle separation.

[0025] A “test sample” comprises genomic nucleic acids typically from a patient or a cell population suspected of, or being screened for, nucleic acid containing a chromosomal or genetic abnormality. A “reference sample” comprises genomic nucleic acids typically from a normal individual or wild-type cell population with a normal genetic profile. A “test sample” also may comprise mRNA from which cDNA can be made.

[0026] The genomic nucleic acids from the test and reference samples are contacted under hybridization conditions to a surface containing a plurality of nucleic acid segments, each immobilized at discrete locations on the surface. The term “hybridization” as used herein, refers to the pairing of substantially complementary nucleotide sequences (strands of nucleic acid) to form a duplex or heteroduplex through formation of hydrogen bonds between complementary base pairs. It is a specific, *i.e.*, non-random, interaction between two complementary polynucleotides. Hybridization and the strength of hybridization (*i.e.*, the strength of the association between the nucleic acids) is influenced by such factors as the degree of complementary between the nucleic acids, stringency of the conditions involved, and the T_m of the formed hybrid.

[0027] Genomic nucleic acids of the test sample may be linked to a detectable label via a first linkage. Genomic nucleic acids of the reference sample may be linked to the same detectable label via a second linkage.

[0028] A “detectable label” as used herein refers any moiety that generates a detectable signal by spectroscopic, photochemical, biochemical, immunochemical, electromagnetic, radiochemical, or chemical means, such as fluorescence, chemifluorescence, or chemiluminescence, or any other appropriate means. Preferred detectable labels include fluorescent dye molecules, or fluorophores, such as fluorescein, phycoerythrin, Cy3TM, Cy5TM, allophycocyanine, Texas Red, peridenin chlorophyll, cyanine, FAM, JOE, TAMRA, tandem conjugates such as phycoerythrin-Cy5TM, and the like. Cy3TM, and Cy5TM are commonly used together in two color detection systems. In single label CGH, Cy3TM is preferred over Cy5TM. The detectable label may be linked by covalent or non-covalent means to nucleic acids. Alternatively, a detectable label may be linked such as by directly labeling a molecule that achieves binding to another nucleic acid via a ligand-receptor binding pair arrangement or other such specific recognition molecules.

[0029] A “linkage” of a detectable label as used herein, means that the label is physically associated with genomic nucleic acids in a sample. In one embodiment, either the first linkage or the second linkage is susceptible to selective removal, *i.e.*, the linkage of one label but not the other is susceptible to cleavage or separation allowing the label to be separated from the nucleic acid. Examples of pairs of linkages (*i.e.*, a differential linkage where one linker of the pair is susceptible to selective removal) include linkage via two different chemical linkers, two different oligonucleotides, or two different peptide sequences, wherein

the chemical linkers, oligonucleotides or peptide sequences differ in susceptibility to temperature, pH hydrolysis, radiation (*e.g.*, nucleotide stretches or chemical entities sensitive to ultraviolet radiation; *e.g.*, photocleavable entities), oxidative conditions, atmospheric conditions (*e.g.*, exposure to ozone), buffer conditions, hydrolysis by an external agent (*e.g.*, an enzyme, such as a restriction endonuclease or a homing endonuclease), or chemical cleavage (*e.g.*, linkers containing a diol that can be selectively cleaved using saturated aqueous NaIO₄ for 30-40 minutes, or linkers containing a disulfide that can be cleaved with dithiothreitol or any other appropriate reducing reagent, such as those available from Fidelity Systems, Inc. Gaithersburg, MD).

[0030] By “susceptibility” is meant that the detectable label associated with the nucleic acids containing the susceptible linkage is physically dissociated from about 80% or more, preferably 90% or more, more preferably 95% or more of the member nucleic acids of that sample. Treatments that remove less than 95% can be tolerated especially when the deficiency in removal can be calculated and factored into the final results. Other differential linkages include different chemical coupling or physical interactions of the label with nucleic acids of either the test or reference sample in the labeling process. The nucleic acids of the sample(s) may be labeled before hybridization to the array, or after hybridization to the array. In all of these examples, a susceptible linkage is created to render the nucleic acids of one sample subject to selective removal of the label associated with the nucleic acids in that sample, following an initial read of the hybridization. Thus, the same label may be read at two time points: (i) to detect the signal from the detectable label of both the test and reference nucleic acids hybridized to the array; and (ii) to detect the signal from the nucleic acids hybridized to the array that do have the label attached by a linker that is susceptible to the treatment. Subtraction of these two readings yields a value representing the hybridized nucleic acid that is labeled via the linkage that was resistant to the removal treatment.

[0031] In a one approach, the label may be non-covalently associated with the sample or reference nucleic acid, thus allowing the nucleic acids to first be hybridized to the target before the label is attached. In this case, a first unique oligonucleotide can be attached to either the test or reference sample nucleic acids, wherein this oligonucleotide contains a unique hybridization sequence and a recognition site for a restriction endonuclease, a homing endonuclease, or a rare-cutting endonuclease as is known in the art and commercially available, for example from Fermentas Life Sciences (*e.g.*, I-SceI). The other of the nucleic acids is preferably linked to a second oligonucleotide that contains a different unique

hybridization sequence but does not contain this recognition site. Both nucleic acids with the attached oligonucleotides are separately detected using another detectably labeled oligonucleotide which is complementary to one or the other hybridization sequence. Such detectably labeled complementary oligonucleotide may be a dendrimeric complex. Using this embodiment, following application of both labeled oligonucleotides to the hybridized array, label associated via the first unique oligonucleotide may then be selectively removed by contact with the endonuclease specific for the recognition site in the first unique oligonucleotide.

[0032] A “dendrimer” as used herein, is an artificially manufactured or synthesized polymeric molecule built up from branched units called monomers. In a preferred embodiment, the monomers are DNA molecules which associate by base pairing to assemble (see, *e.g.*, U.S. Patent Nos. 5,175,270; 5,484,904; and 5,487,973). Other monomers include, but are not limited to, primary amines (see, *e.g.*, U.S. Patent No. 5,530,092); polyamidoamines, polyethyleneimines, and polypropyleneimines (see, *e.g.*, U.S. Patent Nos. 5,393,797; 5,393,795; 5,560,929; and 5,387,617); peptides; and other nucleic acids. Various tags (*e.g.*, an oligonucleotide) may be attached to a terminal end of a dendrimer polymer or may be incorporated into the internal structure of the dendrimer. Attachment includes covalent attachment (*e.g.*, the 3' end of an oligonucleotide is covalently attached to a terminal end of a dendrimer branch) as well as non-covalent interactions (*e.g.*, nucleic acid hybridization). By incorporating a multiplicity of labels into the dendrimer, hybridization signal intensity is dramatically enhanced.

[0033] According to yet a further aspect of the invention, there is provided a method of comparing the expression of genes in a test sample versus that of reference sample. The first step of the method includes contacting under hybridization conditions cDNA prepared from mRNA of a test sample and cDNA prepared from mRNA of a reference sample to a surface containing a plurality of nucleic acid segments each immobilized at discrete locations on the surface. In this case, the test sample cDNA and the reference sample cDNA are labeled before or after hybridization with the same detectable label which is linked to the cDNA of the test sample via a first linkage, and to the cDNA of the reference sample via a second linkage. Either the first linkage or the second linkage is susceptible to selective removal and the detectable label linked to nucleic acids hybridized to the surface determined. The location and amount of detectable label linked to nucleic acids hybridized the surface of the support is determined. The label is then selectively removed from either the hybridized test

sample cDNA or the hybridized reference sample cDNA. The location and amount of the detectable label remaining on the support is then determined and represents one of the samples. The difference between the location and amount remaining after removal compared and the location and amount prior to removal represents the other of the samples. The relative amount of each sample nucleic acid hybridized to the array reflects the expression of genes in the test sample compared to the reference sample.

[0034] The term “expression array” refers to a collection of cDNA sequences representing the complement of mRNA present in a cell at a particular time. A cDNA expression array may be prepared by oligo dT priming or random priming with hexomers. The random primer oligos may have 5’ ligatable ends. After priming and extension to produce cDNA, the ligatable ends of the cDNA are ligated to a capture sequence via a bridging oligo. cDNA with capture sequences may then be hybridized to a solid phase containing immobilized nucleic acid probe sequences. The capture sequence may be detected with an appropriate labeled reagent (e.g. labeled dendrimeric nucleic acid) with a single stranded segment(s) (a.k.a. tail(s)) for the capture sequence. Exemplary supplies and protocols for preparing an expression array are available from manufacturers, for example, Genisphere, Inc.

[0035] Any of the embodiments disclosed herein for single label comparative genomic hybridization may be used to detect a chromosomal abnormality or genetic abnormality in any patient including adults, children and neonates.

BRIEF DESCRIPTION OF THE FIGURES

[0036] FIG. 1 illustrates an exemplary embodiment of the present invention. The test and reference genomic DNA fragments are labeled with the same fluorescent label, wherein the label is selectively removable from the test DNA fragments by enzymatic cleavage. The microarray is scanned at two time points, before and after selective removal of the label, thus providing a subtractive approach that allows the use of the same uniform label for CGH comparisons.

[0037] FIG. 2 illustrates an alternative exemplary embodiment of the present invention. The test and reference genomic DNA fragments are sequentially labeled with the same fluorescent label. The microarray is scanned at two time points, after label incorporation into the reference genomic DNA fragments, and after label incorporation into the test genomic

DNA fragments, thus providing an additive approach that allows the use of the same uniform label for CGH comparisons.

[0038] FIG. 3 illustrates an alternative exemplary embodiment of the present invention. The test and reference genomic DNA fragments labeled with distinguishable tag sequences (triple hash marks) are hybridized unlabeled to a duplicate set of arrays overnight (o/n) under a single coverslip. One array is hybridized separately to a labeled dendrimer with an entity specific for the test sample tag. After washing and scanning, both arrays are simultaneously hybridized to a second labeled dendrimer with an entity specific for the reference nucleic acid tag. After washing, a second scan is obtained. The two scans are used as indicated to determine the relative amounts of test and reference genomic DNA that hybridized, thus providing an approach to detect chromosomal or genetic abnormalities.

[0039] FIG. 4 illustrates an alternative exemplary embodiment of the present invention. The test and reference genomic DNA fragments labeled with distinguishable tag sequences (triple hash marks) are hybridized unlabeled together with a labeled dendrimer containing an entity specific for the test nucleic acid to a duplicate set of arrays overnight (o/n) under a single coverslip. After washing and scanning, both arrays are co-hybridized to a labeled dendrimer with an entity specific for the reference sample tag. The arrays are washed and again and scanned. The two scans are used as indicated to determine the relative amounts of test and reference genomic DNA that hybridized, thus providing an approach to detect chromosomal or genetic abnormalities.

DETAILED DESCRIPTION OF THE INVENTION

[0040] In accordance with the present invention, there are provided methods of performing array-based comparative genomic hybridization (CGH) to detect a chromosomal abnormality in a test sample, or to diagnose a genetic abnormality in an individual. CGH is a molecular cytogenetics approach that can be used to detect regions in a genome undergoing quantitative changes, *e.g.*, gains or losses of sequence or copy numbers. CGH is especially useful in the analysis and diagnosis of cancer, and the analysis and diagnosis of genetic disorders, such as in prenatal diagnosis. CGH reactions are typically used to compare the genetic composition of an unknown test sample with a known normal reference sample.

[0041] Sources of Genomic Nucleic Acids

[0042] In one aspect, the methods of the present invention can be used to detect a chromosomal abnormality in a test sample. In a preferred embodiment, the test sample is obtained from a patient. In another preferred embodiment, the test sample contains cells, tissues or fluid obtained from a patient suspected of having a pathology or a condition associated with a chromosomal or genetic abnormality. The causality, diagnosis or prognosis of the pathology or condition may be associated with genetic defects, *e.g.*, with genomic nucleic acid base substitutions, amplifications, deletions and/or translocations. The test sample may be suspected of containing cancerous cells or nucleic acid from such cells. Samples may include, but are not limited to, amniotic fluid, biopsies, blood, blood cells, bone marrow, cerebrospinal fluid, fecal samples, fine needle biopsy samples, peritoneal fluid, plasma, pleural fluid, saliva, semen, serum, sputum, tears, tissue or tissue homogenates, tissue culture media, urine, and the like. Samples may also be processed, such as sectioning of tissues, fractionation, purification, or cellular organelle separation.

[0043] Methods of isolating cell, tissue or fluid samples are well known to those of skill in the art and include, but are not limited to, aspirations, tissue sections, drawing of blood or other fluids, surgical or needle biopsies, and the like. Samples derived from a patient may include frozen sections or paraffin sections taken for histological purposes. The sample can also be derived from supernatants (of cell cultures), lysates of cells, cells from tissue culture in which it may be desirable to detect levels of mosaicisms, including chromosomal abnormalities and copy numbers.

[0044] In a preferred embodiment, a sample suspected of containing cancerous cells is obtained from a human patient. Samples can be derived from patients using well-known techniques such as venipuncture, lumbar puncture, fluid sample such as saliva or urine, tissue or needle biopsy, and the like. In a patient suspected of having a tumor containing cancerous cells, a sample may include a biopsy or surgical specimen of the tumor, including for example, a tumor biopsy, a fine needle aspirate, or a section from a resected tumor. A lavage specimen may be prepared from any region of interest with a saline wash, for example, cervix, bronchi, bladder, *etc.* A patient sample may also include exhaled air samples as taken with a breathalyzer or from a cough or sneeze. A biological sample may also be obtained from a cell or blood bank where tissue and/or blood are stored, or from an *in vitro* source,

such as a culture of cells. Techniques for establishing a culture of cells for use as a sample source are well known to those of skill in the art.

[0045] In another aspect, the methods of the present invention can be used to detect a chromosomal or genetic abnormality in a fetus. Prenatal diagnosis of a fetus may be indicated for women at increased risk of carrying a fetus with chromosomal or genetic abnormalities. Risk factors are well known in the art, and include, for example, advanced maternal age, abnormal maternal serum markers in prenatal screening, chromosomal abnormalities in a previous child, a previous child with physical anomalies and unknown chromosomal status, parental chromosomal abnormality, and recurrent spontaneous abortions.

[0046] The invention methods can be used to perform prenatal diagnosis using any type of embryonic or fetal cell. Fetal cells can be obtained through the pregnant female, or from a sample of an embryo. Thus, fetal cells are present in amniotic fluid obtained by amniocentesis, chorionic villi aspirated by syringe, percutaneous umbilical blood, a fetal skin biopsy, a blastomere from a four-cell to eight-cell stage embryo (pre-implantation), or a trophoctoderm sample from a blastocyst (pre-implantation or by uterine lavage). Body fluids with sufficient amounts of genomic nucleic acid also may be used.

[0047] The method of the present invention utilizes a first population of genomic nucleic acids obtained from the test sample, and a second population of genomic nucleic acids obtained from a reference sample. The reference sample may be any cells, tissues or fluid as provided herein, obtained from an individual, or any cell culture or tissue culture, that does not contain any genetic abnormality, *i.e.*, that has a normal genetic complement of all chromosomes.

[0048] Association of Label with Genomic Nucleic Acids

[0049] The genomic nucleic acids of both the test sample and the reference sample are associated with the same detectable label, either prior to or subsequent to hybridization. In preferred embodiments, the label is detectable by optical means, and is most preferably a fluorescent label or fluorophore. The detectable label can be incorporated into, associated with or conjugated to a nucleic acid. The association between the nucleic acid and the detectable label can be covalent or non-covalent. According to the methods of the present invention, the same detectable label is used to label both the genomic nucleic acids of the test

sample and the genomic nucleic acids of the reference sample. Label can be attached by spacer arms of various lengths to reduce potential steric hindrance or impact on other useful or desired properties. See, *e.g.*, Mansfield, *Mol. Cell. Probes* 9:145-156, 1995.

[0050] Useful labels include, *e.g.*, fluorescent dyes (*e.g.*, Cy5TM, Cy3TM, FITC, rhodamine, lanthanide phosphors, Texas red), ³²P, ³⁵S, ³H, ¹⁴C, ¹²⁵I, ¹³¹I, electron-dense reagents (*e.g.*, gold), enzymes, *e.g.*, as commonly used in an ELISA (*e.g.*, horseradish peroxidase, beta-galactosidase, luciferase, alkaline phosphatase), colorimetric labels (*e.g.*, colloidal gold), magnetic labels (*e.g.*, DynabeadsTM), biotin, dioxigenin, or haptens and proteins for which antisera or monoclonal antibodies are available. The label can be directly incorporated into the nucleic acid to be detected, or it can be attached to a probe (*e.g.*, an oligonucleotide) or antibody that hybridizes or binds to the nucleic acid to be detected.

[0051] In preferred embodiments, the detectable label is a fluorophore. The term “fluorophore” as used herein refers to a molecule that absorbs a quantum of electromagnetic radiation at one wavelength, and emits one or more photons at a different, typically longer, wavelength in response. Suitable fluorescent moieties include the following fluorophores known in the art:

4-acetamido-4'-isothiocyanatostilbene-2,2'-disulfonic acid

acridine and derivatives:

acridine

acridine isothiocyanate

Alexa Fluor[®] 350, Alexa Fluor[®] 488, Alexa Fluor[®] 546, Alexa Fluor[®] 555, Alexa Fluor[®] 568, Alexa Fluor[®] 594, Alexa Fluor[®] 647 (Molecular Probes)

5-(2'-aminoethyl)aminonaphthalene-1-sulfonic acid (EDANS)

4-amino-N-[3-vinylsulfonyl]phenyl]naphthalimide-3,5 disulfonate (Lucifer Yellow VS)

N-(4-anilino-1-naphthyl)maleimide

anthranilamide

Black Hole QuencherTM (BHQTM) dyes (biosearch Technologies)

BODIPY[®] R-6G, BOPIPY[®] 530/550, BODIPY[®] FL

Brilliant Yellow

coumarin and derivatives:

coumarin

7-amino-4-methylcoumarin (AMC, Coumarin 120)

7-amino-4-trifluoromethylcoumarin (Coumarin 151)

Cy2[®], Cy3[®], Cy3.5[®], Cy5[®], Cy5.5[®]

cyanosine

4',6-diaminidino-2-phenylindole (DAPI)

5', 5''-dibromopyrogallol-sulfonephthalein (Bromopyrogallol Red)

7-diethylamino-3-(4'-isothiocyanatophenyl)-4-methylcoumarin

diethylenetriamine pentaacetate

4,4'-diisothiocyanatodihydro-stilbene-2,2'-disulfonic acid

4,4'-diisothiocyanatostilbene-2,2'-disulfonic acid

5-[dimethylamino]naphthalene-1-sulfonyl chloride (DNS, dansyl chloride)

4-(4'-dimethylaminophenylazo)benzoic acid (DABCYL)

4-dimethylaminophenylazophenyl-4'-isothiocyanate (DABITC)

Eclipse[™] (Epoch Biosciences Inc.)

eosin and derivatives:

eosin

eosin isothiocyanate

erythrosin and derivatives:

erythrosin B

erythrosin isothiocyanate

ethidium

fluorescein and derivatives:

5-carboxyfluorescein (FAM)

5-(4,6-dichlorotriazin-2-yl)aminofluorescein (DTAF)

2',7'-dimethoxy-4'5'-dichloro-6-carboxyfluorescein (JOE)

fluorescein

fluorescein isothiocyanate (FITC)

hexachloro-6-carboxyfluorescein (HEX)

QFITC (XRITC)

tetrachlorofluorescein (TET)

fluorescamine

IR144

IR1446

Malachite Green isothiocyanate

4-methylumbelliferone

ortho cresolphthalein

nitrotyrosine

pararosaniline

Phenol Red

B-phycoerythrin, R-phycoerythrin

o-phthaldialdehyde

Oregon Green[®]

propidium iodide

pyrene and derivatives:

pyrene

pyrene butyrate

succinimidyl 1-pyrene butyrate

QSY[®] 7, QSY[®] 9, QSY[®] 21, QSY[®] 35 (Molecular Probes)

Reactive Red 4 (Cibacron[®] Brilliant Red 3B-A)

rhodamine and derivatives:

6-carboxy-X-rhodamine (ROX)

6-carboxyrhodamine (R6G)

lissamine rhodamine B sulfonyl chloride

rhodamine (Rhod)

rhodamine B

rhodamine 123

rhodamine green

rhodamine X isothiocyanate

sulforhodamine B

sulforhodamine 101

sulfonyl chloride derivative of sulforhodamine 101 (Texas Red)

N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA)

tetramethyl rhodamine

tetramethyl rhodamine isothiocyanate (TRITC)

riboflavin

rosolic acid

terbium chelate derivatives

[0052] Other fluorescent nucleotide analogs can be used, see, *e.g.*, Jameson, *Meth. Enzymol.* **278**:363-390, 1997; Zhu, *Nucl. Acids Res.* **22**:3418-3422, 1994. U.S. Patent Nos. 5,652,099 and 6,268,132 also describe nucleoside analogs for incorporation into nucleic acids, *e.g.*, DNA and/or RNA, or oligonucleotides, via either enzymatic or chemical synthesis to produce fluorescent oligonucleotides. U.S. Patent No. 5,135,717 describes phthalocyanine and tetrabenztriazaporphyrin reagents for use as fluorescent labels.

[0053] Detectable labels can be incorporated into nucleic acids by covalent or non-covalent means, *e.g.*, by transcription, such as by random-primer labeling using Klenow polymerase, or nick translation, or, amplification, or equivalent as is known in the art. For example, in one aspect, a nucleoside base is conjugated to a detectable moiety, such as a fluorescent dye, *e.g.*, Cy3TM or Cy5TM, and then incorporated into genomic nucleic acids. Nucleic acids can be incorporated with Cy3TM- or Cy5TM-dCTP conjugates mixed with unlabeled dCTP.

[0054] In another aspect, when using PCR or nick translation to label nucleic acids, modified nucleotides synthesized by coupling allylamine-dUTP to the succinimidyl-ester derivatives of the fluorescent dyes or haptens (such as biotin or digoxigenin) can be used; this method allows custom preparation of most common fluorescent nucleotides, see, *e.g.*, Henegariu, *Nat. Biotechnol.* **18**:345-348, 2000.

[0055] Alternative non-covalent incorporation of label can be achieved using other methods known in the art. For example, Kreatech Biotechnology's Universal Linkage System® (ULS®) provides a non-enzymatic labeling technology, wherein a platinum group forms a co-ordinative bond with DNA, RNA or nucleotides by binding to the N7 position of guanosine. This technology may also be used to label proteins by binding to nitrogen and sulphur containing side chains of amino acids. See, *e.g.*, U.S. Patent Nos. 5,580,990; 5,714,327; and 5,985,566; and European Patent No. 0539466. Thus, this system provides a method of associating any detectable label with members of a nucleic acid population, either directly into a nucleic acid or peptide molecule associated thereto, or indirectly via a complementary nucleic acid molecule or other partner molecule.

[0056] Labeling with a detectable label also can include a nucleic acid attached to another biological molecule, such as a nucleic acid, *e.g.*, an oligonucleotide, or a nucleic acid in the form of a stem-loop structure as a "molecular beacon" or an "aptamer beacon". Molecular beacons as detectable moieties are well known in the art; for example, Sokol (*Proc. Natl.*

Acad. Sci. USA **95**:11538-11543, 1998) synthesized “molecular beacon” reporter oligodeoxynucleotides with matched fluorescent donor and acceptor chromophores on their 5' and 3' ends. In the absence of a complementary nucleic acid strand, the molecular beacon remains in a stem-loop conformation where fluorescence resonance energy transfer prevents signal emission. On hybridization with a complementary sequence, the stem-loop structure opens increasing the physical distance between the donor and acceptor moieties thereby reducing fluorescence resonance energy transfer and allowing a detectable signal to be emitted when the beacon is excited by light of the appropriate wavelength. See also, *e.g.*, Antony (*Biochemistry* **40**:9387-9395, 2001), describing a molecular beacon comprised of a G-rich 18-mer triplex forming oligodeoxyribonucleotide. See also U.S. Patent Nos. 6,277,581 and 6,235,504.

[0057] Aptamer beacons are similar to molecular beacons; see, *e.g.*, Hamaguchi, *Anal. Biochem.* **294**:126-131, 2001; Poddar, *Mol. Cell. Probes* **15**:161-167, 2001; Kaboev, *Nucl. Acids Res.* **28**:E94, 2000. Aptamer beacons can adopt two or more conformations, one of which allows ligand binding. A fluorescence-quenching pair is used to report changes in conformation induced by ligand binding. See also, *e.g.*, Yamamoto, *Genes Cells* **5**:389-396, 2000; Smimov, *Biochemistry* **39**:1462-1468, 2000.

[0058] In a preferred embodiment, genomic nucleic acids are labeled using an oligonucleotide linkage. The genomic nucleic acids are first digested into fragments with a restriction enzyme (*e.g.*, AluI); fragments are then associated with a unique capture sequence using a bridging oligonucleotide. When properly designed, the unique fragment is positioned directly adjoining the end of a nucleic acid such that ligation can be used to obtain covalent linkage. Each fragment can then be labeled with a dendrimeric construct comprising an oligonucleotide which hybridizes to the unique capture sequence associated with each fragment. The fragments of two or more samples of nucleic acids can be labeled via a unique capture sequence associated with each respective sample. In an especially preferred embodiment, multiple copies of the detectable label are attached to a dendrimer to achieve signal amplification. Preferably, the use of a dendrimer in the methods of the present invention allows more than 10, 20, 50, 100, or 200 fluorophore molecules to be attached to the genomic acids. Labeling of the fragments can be prior to hybridization of two or more nucleic acid samples, or preferably following hybridization to maximize signal intensity.

[0059] Alternatively, the genomic nucleic acid may be labeled via a peptide. A peptide can be made detectable by incorporating predetermined polypeptide epitopes recognized by a secondary reporter (*e.g.*, leucine zipper pair sequences, binding sites for secondary antibodies, transcriptional activator polypeptide, metal binding domains, epitope tags). A label may also be attached via a second peptide (such as on a dendrimer construct as above) that interacts with the first peptide (*e.g.*, S-S association).

[0060] In another embodiment, the genomic nucleic acid may be labeled via a peptide nucleic acid. The term “peptide nucleic acid” (or PNA) as used herein refers to a molecule comprising bases or base analogs such as would be found in natural nucleic acid, but attached to a peptide backbone rather than the sugar-phosphate backbone typical of nucleic acids. The attachment of the bases to the peptide is such as to allow the bases to base pair with complementary bases of nucleic acid in a manner similar to that of an oligonucleotide. These small molecules, also designated anti gene agents, stop transcript elongation by binding to their complementary strand of nucleic acid (Nielsen *et al.*, *Anticancer Drug Des.* 8:53-63, 1993).

[0061] Indirect labeling may be performed prior to or preferably, after hybridization to maximize signal intensity. In a preferred embodiment, the hybridized surface is contacted with a first complex containing a detectable label and a first entity, wherein the first complex selectively reacts with the nucleic acids of either the test sample or the reference sample; and either simultaneously or subsequently with a second complex containing the same detectable label and a second entity, wherein the second complex selectively reacts with the nucleic acids of the other sample. In one embodiment, the first complex or the second complex may comprise a differential linkage of the detectable label, such that one sample may be subjected to selective removal of the detectable label (*i.e.*, a subtractive approach). Alternatively, in another embodiment, the first complex and the second complex do not comprise a differential linkage of the detectable label, but instead, are added one following the other (*i.e.*, an additive approach).

[0062] In certain embodiments, isolated or purified molecules may be preferred. As used herein, the terms “isolated”, “purified” or “substantially purified” refer to molecules, either nucleic acid or amino acid sequences, that are removed from their natural environment, isolated or separated, and are at least 60% free, preferably 75% free, and most preferably

90% free from other components with which they are naturally associated. An isolated molecule is therefore a substantially purified molecule.

[0063] Hybridization

[0064] The methods of the present invention can incorporate all known methods and means and variations thereof for carrying out comparative genomic hybridization, see, *e.g.*, U.S. Patent Nos. 6,197,501; 6,159,685; 5,976,790; 5,965,362; 5,856,097; 5,830,645; 5,721,098; 5,665,549; 5,635,351; Diago, *Am. J. Pathol.* **158**:1623-1631, 2001; Theillet, *Bull. Cancer* **88**:261-268, 2001; Werner, *Pharmacogenomics* **2**:25-36, 2001; Jain, *Pharmacogenomics* **1**:289-307, 2000.

[0065] Generally, nucleic acid hybridizations comprise the following major steps: (1) immobilization of target nucleic acids; (2) pre-hybridization treatment to increase accessibility of target DNA, and to reduce nonspecific binding; (3) hybridization of the mixture of nucleic acids to the nucleic acid on the solid surface; (4) post-hybridization washes to remove nucleic acid fragments not bound in the hybridization and (5) detection of the hybridized nucleic acid fragments. If indirect detection is used, an additional step of hybridization with the labeled agent (*e.g.* dendrimer) and washing is needed. The reagent used in each of these steps and their conditions for use vary depending on the particular application.

[0066] In some applications it is necessary to block the hybridization capacity of repetitive sequences. A number of methods for removing and/or disabling the hybridization capacity of repetitive sequences are known (see, *e.g.*, WO 93/18186). For instance, bulk procedures can be used. In many genomes, including the human genome, a major portion of shared repetitive DNA is contained within a few families of highly repeated sequences such as Alu. These methods exploit the fact that hybridization rate of complementary sequences increases as their concentration increases. Thus, repetitive sequences, which are generally present at high concentration will become double stranded more rapidly than others following denaturation and incubation under hybridization conditions. The double stranded nucleic acids are then removed and the remainder used in hybridizations. Methods of separating single from double stranded sequences include using hydroxyapatite or immobilized complementary nucleic acids attached to a solid support, and the like. Alternatively, the partially hybridized mixture can be used and the double stranded sequences will be unable to hybridize to the target.

[0067] Alternatively, unlabeled sequences which are complementary to the sequences whose hybridization capacity is to be inhibited can be added to the hybridization mixture. This method can be used to inhibit hybridization of repetitive sequences as well as other sequences. For example, Cot-1 DNA can be used to selectively inhibit hybridization of repetitive sequences in a sample. To prepare Cot-1 DNA, DNA is extracted, sheared, denatured and renatured. Because highly repetitive sequences reanneal more quickly, the resulting hybrids are highly enriched for these sequences. The remaining single stranded (*i.e.*, single copy sequences) is digested with S1 nuclease and the double stranded Cot-1 DNA is purified and used to block hybridization of repetitive sequences in a sample. Although Cot-1 DNA can be prepared as described above, it is also commercially available (BRL).

[0068] Hybridization conditions for nucleic acids in the methods of the present invention are well known in the art. Hybridization conditions may be high, moderate or low stringency conditions. Ideally, nucleic acids will hybridize only to complementary nucleic acids and will not hybridize to other non-complementary nucleic acids in the sample. The hybridization conditions can be varied to alter the degree of stringency in the hybridization and reduce background signals as is known in the art. For example, if the hybridization conditions are high stringency conditions, a nucleic acid will bind only to nucleic acid target sequences with a very high degree of complementarity. Low stringency hybridization conditions will allow for hybridization of sequences with some degree of sequence divergence. The hybridization conditions will vary depending on the biological sample, and the type and sequence of nucleic acids. One skilled in the art will know how to optimize the hybridization conditions to practice the methods of the present invention.

[0069] As used herein the term “stringency” is used in reference to the conditions of temperature, ionic strength, and the presence of other compounds, under which nucleic acid hybridizations are conducted. With high stringency conditions, nucleic acid base pairing will occur only between nucleic acids that have sufficiently long segment with a high frequency of complementary base sequences.

[0070] Exemplary hybridization conditions are as follows. High stringency generally refers to conditions that permit hybridization of only those nucleic acid sequences that form stable hybrids in 0.018M NaCl at 65 °C. High stringency conditions can be provided, for example, by hybridization in 50% formamide, 5X Denhardt's solution, 5X SSC (saline sodium citrate) 0.2% SDS (sodium dodecyl sulphate) at 42 °C, followed by washing in 0.1X

SSC, and 0.1% SDS at 65 °C. Moderate stringency refers to conditions equivalent to hybridization in 50% formamide, 5X Denhardt's solution, 5X SSC, 0.2% SDS at 42 °C, followed by washing in 0.2X SSC, 0.2% SDS, at 65 °C. Low stringency refers to conditions equivalent to hybridization in 10% formamide, 5X Denhardt's solution, 6X SSC, 0.2% SDS, followed by washing in 1X SSC, 0.2% SDS, at 50 °C.

[0071] As used herein, the terms “complementary” or “complementarity” are used in reference to polynucleotides (*i.e.*, a sequence of nucleotides such as an oligonucleotide or a target nucleic acid) related by the base-pairing rules. The complement of a nucleic acid sequence as used herein refers to an oligonucleotide which, when aligned with the nucleic acid sequence such that the 5' end of one sequence is paired with the 3' end of the other, is in “antiparallel association.” For example, the sequence “5'-A-G-T-3'” is complementary to the sequence “3'-T-C-A-5'”. Certain bases not commonly found in natural nucleic acids may be included in the nucleic acids of the present invention and include, for example, inosine and 7-deazaguanine. Complementarity need not be perfect; stable duplexes may contain mismatched base pairs or unmatched bases. Those skilled in the art of nucleic acid technology can determine duplex stability empirically considering a number of variables including, for example, the length of the oligonucleotide, base composition and sequence of the oligonucleotide, ionic strength and incidence of mismatched base pairs.

[0072] Complementarity may be “partial” in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be “complete” or “total” complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods that depend upon binding between nucleic acids. Either term may also be used in reference to individual nucleotides, especially within the context of polynucleotides. For example, a particular nucleotide within an oligonucleotide may be noted for its complementarity, or lack thereof, to a nucleotide within another nucleic acid strand, in contrast or comparison to the complementarity between the rest of the oligonucleotide and the nucleic acid strand.

[0073] The term “homology” and “homologous” refers to a degree of identity between two sequences. There may be partial homology or complete homology. A partially homologous sequence is one that is less than 100% identical to another sequence. Preferably,

homologous sequences have an overall identity of at least 70% or at least 75%, more preferably at least 80% or at least 85%, most preferably at least 90% or at least 95%.

[0074] As used herein, the term “ T_m ” is used in reference to the “melting temperature”. The melting temperature is the temperature at which a sample of double-stranded nucleic acid molecules becomes half dissociated into single strands. Several equations for calculating the T_m of nucleic acids are well known in the art. As indicated by standard references, a simple estimate of the T_m value may be calculated by the equation: $T_m = 81.5 + 0.41(\% G+C)$, when a nucleic acid is in aqueous solution at 1 M NaCl (see, *e.g.*, Anderson and Young, Quantitative Filter Hybridization, in Nucleic Acid Hybridization, 1985). Other references (*e.g.*, Allawi and SantaLucia, *Biochemistry* 36:10581-94, 1997) include more sophisticated computations which take structural and environmental, as well as sequence characteristics into account for the calculation of T_m .

[0075] Arrays

[0076] Nucleic acids used in the methods of the present invention can be immobilized to or applied to an array or “biochip”. The term “array” or “microarray” or “biochip” or “chip” as used herein refers to a plurality of elements arranged onto a defined area of a substrate surface. In practicing the methods of the invention, any known array and/or method of making and using arrays can be incorporated in whole or in part, or variations thereof, as disclosed, for example, in U.S. Patent Nos. 6,277,628; 6,277,489; 6,261,776; 6,258,606; 6,054,270; 6,048,695; 6,045,996; 6,022,963; 6,013,440; 5,965,452; 5,959,098; 5,856,174; 5,830,645; 5,770,456; 5,632,957; 5,556,752; 5,143,854; 5,807,522; 5,800,992; 5,744,305; 5,700,637; 5,556,752; 5,434,049; see also, *e.g.*, WO 99/51773; WO 99/09217; WO 97/46313; WO 96/17958; see also, *e.g.*, Johnston, *Curr. Biol.* 8:R171-R174, 1998; Schummer, *Biotechniques* 23:1087-1092, 1997; Kern, *Biotechniques* 23:120-124, 1997; Solinas-Toldo, *Genes, Chromosomes & Cancer* 20:399-407, 1997; Bowtell, *Nature Genetics Supp.* 21:25-32, 1999. See also published U.S. Patent Applications Nos. 20010018642; 20010019827; 20010016322; 20010014449; 20010014448; 20010012537; 20010008765.

[0077] Arrays are generically a plurality of “target elements” or “spots,” each target element containing a defined amount of one or more biological molecules, *e.g.*, polypeptides, nucleic acid molecules, or probes, immobilized at discrete locations on a substrate surface. In preferred embodiments, the plurality of spots comprises nucleic acid segments, immobilized at preferably at least about 50, at least about 100, at least about 300, or at least about 500

discrete locations on the surface. The plurality may comprise multiple repeats of the same nucleic acid segments to produce, *e.g.*, duplicate spots, triplicate spots, quadruplicate spots, quintuplicate spots, *etc.*

[0078] The resolution of array-based CGH is primarily dependent upon the number, size and map positions of the nucleic acid elements within the array, which are capable of spanning the entire genome. Each nucleic acid of interest to be immobilized may be contained within a nucleic acid vector (*e.g.*, plasmids, cosmids, *etc.*), or an artificial chromosome, such as a bacterial artificial chromosome (BAC) or P-1 derived artificial chromosome as is known in the art, which are capable of incorporating large inserts of nucleic acid. Typically, bacterial artificial chromosomes, or BACs, which can each accommodate on average about 150 kilobases (kb) of cloned genomic DNA, are used in the production of the array. Preferably, each nucleic acid segment of interest is between about 1,000 (1 kB) and about 1,000,000 (1 MB) nucleotides in length, more preferably between about 100,000 (100kB) and 300,000 (kB) nucleotides in length. Nucleic acid segments of interest may be chosen to span (*i.e.* collectively represent) the sequence of at least one chromosome, spaced at intervals along the chromosome (*i.e.* containing segments of chromosomal sequence) of about 3-4 megabases (MB), more preferably at intervals of about 2-3 megabases along the chromosome, most preferably at intervals of about 1-2 megabases along the chromosome. To represent the entire genomic complement, nucleic acid segments may be chosen to span all chromosomes at such intervals. Alternatively, selected genomic regions of interest, *e.g.*, known mutational hotspots, may be selected from one or more chromosomes. Such genomic regions of interest may be nucleic acid segments associated with a chromosomal abnormality, a contiguous gene abnormality, a genetically linked disease or syndrome.

[0079] Typically, the immobilized nucleic acid molecules are contacted with a sample for specific binding, *e.g.*, hybridization, between molecules in the sample and the array. Immobilized nucleic acids segments can contain sequences from specific messages (*e.g.*, as cDNA libraries) or genes (*e.g.*, genomic libraries), including, *e.g.*, substantially all or a subsection of a chromosome or substantially all of a genome, including a human genome. Other target elements can contain reference sequences, such as positive and negative controls, and the like. The target elements of the arrays may be arranged on the substrate surface at different sizes and different densities. Different target elements of the arrays can have the same molecular species, but, at different amounts, densities, sizes, labeled or unlabeled, and

the like. The target element sizes and densities will depend upon a number of factors, such as the nature of the label (the immobilized molecule can also be labeled), the substrate support (it is solid, semi-solid, fibrous, capillary or porous), and the like.

[0080] Each target element may comprise substantially the same nucleic acid sequences, or, a mixture of nucleic acids of different lengths and/or sequences. Thus, for example, a target element may contain more than one copy of a cloned piece of DNA, and each copy may be broken into fragments of different lengths, as described herein. The length and complexity of the nucleic acid fixed onto the array surface is not critical to the invention. The array can comprise nucleic acids immobilized on any substrate, *e.g.*, a solid surface (*e.g.*, nitrocellulose, glass, quartz, fused silica, plastics and the like). See, *e.g.*, U.S. Patent No. 6,063,338 describing multi-well platforms containing cycloolefin polymers if fluorescence is to be measured. Arrays used in the methods of the invention can comprise housing containing components for controlling humidity and temperature during the hybridization and wash reactions.

[0081] The CGH methods of the invention can be performed using any type of array. Commercially available CGH arrays or prepared slides for array printing include, for example, GeneChipsTM from Affymetrix, Santa Clara, Calif.; Spectral chipTM Mouse BAC Arrays and Spectral chipTM Human BAC Arrays and other custom Arrays from Spectral Genomics, Houston, Tex.; codelinkTM Human Bioarrays from Amersham Biosciences (GE Healthcare); and UltraGapTM from Dow Corning, Elizabethtown, Kentucky. UltraGapTM slides used in accordance with the manufacturer's suggested protocol are preferred.

[0082] In a preferred embodiment, the surface comprises an array containing one, several or all of the human genomic nucleic acid segments provided in a compendium of bacterial artificial chromosomes (BACs) compiled by The BAC Resource Consortium, and referred to in the art by their RPI or CTB clone names, see Cheung *et al.*, *Nature* **409**:953-958, 2001. This compendium contains 7,600 cytogenetically defined landmarks on the draft sequence of the human genome (see McPherson *et al.*, *Nature* **409**:934-41, 2001). These landmarks are large-insert clones mapped to chromosome bands by fluorescence *in situ* hybridization, each containing a sequence tag that is positioned on the genomic sequence. These clones represent all 24 human chromosomes in about 1 Mb resolution. Sources of BAC genomic collections include the BACPAC Resources Center (CHORI - Children's Hospital Oakland Research Institute), ResGen (Research Genetics through Invitrogen) and The Sanger Center (UK).

[0083] Many methods for immobilizing nucleic acids on a variety of solid surfaces are known in the art. For instance, the solid surface may be a membrane, glass, plastic, or a bead. The desired component may be covalently bound or noncovalently attached through nonspecific binding. The immobilization of nucleic acids on solid surfaces is discussed more fully below.

[0084] A wide variety of organic and inorganic polymers, as well as other materials, both natural and synthetic, may be employed as the material for the solid surface. Illustrative solid surfaces include nitrocellulose, nylon, glass, diazotized membranes (paper or nylon), silicones, polyformaldehyde, cellulose, and cellulose acetate. In addition, plastics such as polyethylene, polypropylene, polystyrene, and the like can be used. Other materials which may be employed include paper, ceramics, metals, metalloids, semiconductive materials, cermets or the like. In addition substances that form gels can be used. Such materials include proteins (*e.g.*, gelatins), lipopolysaccharides, silicates, agarose and polyacrylamides. Where the solid surface is porous, various pore sizes may be employed depending upon the nature of the system.

[0085] In preparing the surface of a solid support for array printing, a plurality of different materials may be employed, particularly as laminates, to obtain various properties. For example, proteins (*e.g.*, bovine serum albumin) or mixtures of macromolecules (*e.g.*, Denhardt's solution) can be employed to avoid non-specific binding, simplify covalent conjugation, enhance signal detection or the like.

[0086] If covalent bonding between a compound and the surface is desired, the surface may be polyfunctional or be capable of being polyfunctionalized. Functional groups which may be present on the surface and used for linking can include carboxylic acids, aldehydes, amino groups, cyano groups, ethylenic groups, hydroxyl groups, mercapto groups and the like. The manner of linking a wide variety of compounds to various surfaces is well known and is amply illustrated in the literature. For example, methods for immobilizing nucleic acids by introduction of various functional groups to the molecules is known (see, *e.g.*, Bischoff *et al.*, *Anal. Biochem.* **164**:336-344, 1987); Kremsky *et al.*, *Nucl. Acids Res.* **15**:2891-2910, 1987). Modified nucleotides can be placed on the target using PCR primers containing the modified nucleotide, or by enzymatic end labeling with modified nucleotides.

[0087] Alternative surfaces include derivatized surfaces such as chemically coated glass slides. One example, is the CodeLink™ Activated Slide from Amersham Biosciences. These

slides are coated with a novel 3-D surface chemistry comprised of a long-chain, hydrophilic polymer containing amine-reactive groups, to react with and covalently immobilize amine-modified DNA for microarrays. This polymer is covalently crosslinked to itself and to the surface of the slide and is designed to orient the immobilized DNA away from the surface of the slide to improve hybridization. Another such 3D slide is UltraGapTM, sold by Dow Corning.

[0088] Use of membrane supports (e.g., nitrocellulose, nylon, polypropylene) for the nucleic acid arrays of the invention is advantageous because of well developed technology employing manual and robotic methods of arraying targets at relatively high element densities (e.g., up to 30-40/cm²). In addition, such membranes are generally available and protocols and equipment for hybridization to membranes is well known. Many membrane materials, however, have considerable fluorescence emission, where fluorescent labels are used to detect hybridization.

[0089] To optimize a given assay format one of skill can determine sensitivity of fluorescence detection for different combinations of membrane type, fluorophore, excitation and emission bands, spot size and the like. In addition, low fluorescence background membranes have been described (see, e.g., Chu *et al.*, *Electrophoresis* **13**:105-114, 1992).

[0090] The sensitivity for detection of spots of various diameters on the candidate membranes can be readily determined by, for example, spotting a dilution series of fluorescently end labeled DNA fragments. These spots are then imaged using conventional fluorescence microscopy. The sensitivity, linearity, and dynamic range achievable from the various combinations of fluorophore and membranes can thus be determined. Serial dilutions of pairs of fluorophore in known relative proportions can also be analyzed to determine the accuracy with which fluorescence ratio measurements reflect actual fluorophore ratios over the dynamic range permitted by the detectors and membrane fluorescence.

[0091] Arrays on substrates with much lower fluorescence than membranes, such as glass, quartz, or small beads, can achieve much better sensitivity. For example, elements of various sizes, ranging from about 1 mm diameter down to about 1 μ m can be used with these materials. Small array members containing small amounts of concentrated target DNA are conveniently used for high complexity comparative hybridizations since the total amount of probe available for binding to each element will be limited. Thus, it is advantageous to have small array members that contain a small amount of concentrated target DNA so that the

signal that is obtained is highly localized and bright. Such small array members are typically used in arrays with densities greater than $10^4/\text{cm}^2$. Relatively simple approaches capable of quantitative fluorescent imaging of 1 cm^2 areas have been described that permit acquisition of data from a large number of members in a single image (see, *e.g.*, Wittrup *et al.*, *Cytometry* 16:206-213, 1994).

[0092] Covalent attachment of the target nucleic acids to glass or synthetic fused silica can be accomplished according to a number of known techniques. Such substrates provide a very low fluorescence substrate, and a highly efficient hybridization environment.

[0093] There are many possible approaches to coupling nucleic acids to glass that employ commercially available reagents. For instance, materials for preparation of silanized glass with a number of functional groups are commercially available or can be prepared using standard techniques. Alternatively, quartz cover slips, which have at least 10-fold lower auto fluorescence than glass, can be silanized.

[0094] The targets can also be immobilized on commercially available coated beads or other surfaces. For instance, biotin end-labeled nucleic acids can be bound to commercially available avidin-coated beads. Streptavidin or anti-digoxigenin antibody can also be attached to silanized glass slides by protein-mediated coupling, using *e.g.*, protein A following standard protocols (see, *e.g.*, Smith *et al.*, *Science* 258:1122-1126, 1992). Biotin or digoxigenin end-labeled nucleic acids can be prepared according to standard techniques.

[0095] Hybridization to nucleic acids attached to beads is accomplished by suspending them in the hybridization mix, and then depositing them on the glass substrate for analysis after washing. Alternatively, paramagnetic particles, such as ferric oxide particles, with or without avidin coating, can be used.

[0096] Interpretation of Array-based CGH

[0097] The copy number of particular nucleic acid sequences in a test sample and a reference sample are compared by hybridizing the samples to one or more target nucleic acid segments. The hybridization signal intensity, and the ratio of intensities, produced by the detectable label associated with each sample is determined. Typically, the greater the ratio of the signal intensities on a target nucleic acid segment, the greater the copy number ratio of sequences in the two samples that bind to that element. Thus comparison of the signal

intensity ratios among target nucleic acid segments permits comparison of copy number ratios of different sequences in the genomic nucleic acids of the two samples.

[0098] In addition to labeling nucleic acids with fluorescent dyes, the invention can be practiced using any apparatus or methods to detect detectable labels associated with nucleic acids of a sample, an individual member of the nucleic acids of a sample, or an array-immobilized nucleic acid segment, or, any apparatus or methods to detect nucleic acids specifically hybridized to each other. Devices and methods for the detection of multiple fluorophores are well known in the art, see, *e.g.*, U.S. Patent Nos. 5,539,517; 6,049,380; 6,054,279; 6,055,325; and 6,294,331. Any known device or method, or variation thereof, can be used or adapted to practice the methods of the invention, including array reading or “scanning” devices, such as scanning and analyzing multicolor fluorescence images; see, *e.g.*, U.S. Patent Nos. 6,294,331; 6,261,776; 6,252,664; 6,191,425; 6,143,495; 6,140,044; 6,066,459; 5,943,129; 5,922,617; 5,880,473; 5,846,708; 5,790,727; and, the patents cited in the discussion of arrays, herein. See also published U.S. Patent Application Nos. 20010018514; 20010007747; and published international patent applications Nos. WO0146467 A; WO9960163 A; WO0009650 A; WO0026412 A; WO0042222 A; WO0047600 A; and WO0101144 A.

[0099] For example a spectrograph can image an emission spectrum onto a two-dimensional array of light detectors; a full spectrally resolved image of the array is thus obtained. Photophysics of the fluorophore, *e.g.*, fluorescence quantum yield and photodestruction yield, and the sensitivity of the detector are read time parameters for an oligonucleotide array. With sufficient laser power and use of Cy5TM or Cy3TM, which have lower photodestruction yields an array can be read in less than 5 seconds.

[00100] Charge-coupled devices, or CCDs, are used in microarray scanning systems, including practicing the methods of the invention. Color discrimination can also be based on 3-color CCD video images; these can be performed by measuring hue values. Hue values are introduced to specify colors numerically. Calculation is based on intensities of red, green and blue light (RGB) as recorded by the separate channels of the camera. The formulation used for transforming the RGB values into hue, however, simplifies the data and does not make reference to the true physical properties of light. Alternatively, spectral imaging can be used; it analyzes light as the intensity per wavelength, which is the only quantity by which to describe the color of light correctly. In addition, spectral imaging can provide spatial data,

because it contains spectral information for every pixel in the image. Alternatively, a spectral image can be made using brightfield microscopy, see, *e.g.*, U.S. Patent No. 6,294,331.

[00101] A specific advantage of the methods of the present invention is that a single detectable label may be used. This eliminates the need to read and co-ordinate multiple colored fluorophores. Thus, signal intensity at the lower range is uniform and can readily be normalized, as opposed to having to account for differences in signal intensity amongst more than one fluorophore. Other advantages of the present invention's array-based CGH approach include the increased resolution by spanning across the entire genomic sequence of each chromosome and the increased sensitivity achieved as compared to traditional *in situ* chromosomal hybridization.

[00102] The methods of the invention further comprise data analysis, which can include the steps of determining, *e.g.*, fluorescent intensity as a function of substrate position, removing "outliers" (data deviating from a predetermined statistical distribution), or calculating the relative binding affinity of the targets from the remaining data. The resulting data can be displayed as an image with color in each region varying according to the light emission or binding affinity between targets and probes. See, *e.g.*, U.S. Patent Nos. 5,324,633; 5,863,504; and 6,045,996. The invention can also incorporate a device for detecting a labeled marker on a sample located on a support, see, *e.g.*, U.S. Patent No. 5,578,832.

[00103] The invention will now be described in greater detail by reference to the following non-limiting examples.

EXAMPLE 1

Preparation of a Genomic Nucleic Acid Array

[00104] A variety of microarray equipment (*e.g.*, BioRobotics Microgrid and others; collectively "arrayers") are available for printing the nucleic acid material onto a plurality of discrete locations of a solid surface. Two specific surfaces were printed with native BAC DNA to establish a protocol for the specific application of large-insert clone microarray fabrication (*e.g.*, BACs, PACs, cosmids).

[00105] Typical prior art arrayer installation and validation protocols assess the printing performance of an arrayer using either dye-only solutions or dye-oligo DNA solutions. These conditions do not reflect the fluid dynamics associated with large clone array manufacturing

and hence are sub-optimal for generating printing parameters. The present example described herein establishes a simple and qualitative approach to validating arrayers and establishing printing parameters for large insert clone microarray fabrication.

[00106] A sample collection of the large insert DNA clones (BACs, PACs, cosmids) intended for printing was resuspended in a salt containing printing buffer (*e.g.*, 50-150 mM sodium phosphate, pH 8-9) at a concentration of 75-100 ng/ μ l. The DNA was briefly fragmented using an ultrasonic water-bath processor set at 100 A with 70 W output for 5 seconds. Gel electrophoreses (0.8-1.0% agarose) was used to confirm that the size of the fragmented DNA ranged homogenously within 500 base pairs and larger. To a 30 μ l aliquot of the sonicated DNA was added 1 μ l of fluorescent nucleotide dye-conjugate (1 mM) of choice. Samples were mixed and transferred to a printing surface. Upon completion of the printing process, the resulting image was evaluated by scanning with a laser scanner (*e.g.*, Axon 4000, 4100, 4200) set at the wavelength of fluorescent dye used.

[00107] Under these typical parameters, two surfaces were tested. The first surface was plain glass slides cleaned according to a standard base/acid protocol. Fluorescent measurements on plain glass slides indicated a background reading of about 3000, with a spot intensity of about 10,000, and a spot size of approximately 290 μ m. The second surface was the CodeLinkTM Activated Slide (Amersham Biosciences). Fluorescent measurements on the CodeLinkTM Activated Slide indicated a background reading of about 15,000, with a spot intensity of about 65,000, and a spot size of approximately 180 μ m.

EXAMPLE 2

Preparation of Genomic Nucleic Acids

[00108] Labeling. Genomic DNA may be labeled by any standard protocol to incorporate a detectable label. An exemplary random priming with a fluorophore is as follows. In a 100 μ l reaction containing 1 ng to 1 μ g DNA, combine 1x random primers solution (BioPrime DNA Labeling System, Gibco BRL), 1 mM Tris, pH 7.6, 0.1 mM EDTA, 0.2 mM each of dATP, dTTP and dGTP, 0.1 mM dCTP, 0.4 mM Cy3 or Cy5-dCTP (Amersham) and 160 U Klenow fragment (BioPrime DNA Labeling System, Gibco BRL). The DNA and random primers solution is incubated at 100 °C for 10 minutes in a total volume of 84 μ l, prior to adding the other reagents, and then the final 100 μ l reaction is incubated overnight at 37 °C. Unincorporated nucleotides are removed using a Sephadex G-50 column.

[00109] Dendrimeric labeling. Genomic DNA may contain a tag contained within a dendrimeric construct. A dendrimer is a highly branched molecule created to integrate multiple copies of the desired detectable label to amplify detection. Kits for dendrimer labeling and construction are commercially available (*e.g.*, Genisphere Inc.). Briefly, genomic DNA is digested with AluI to yield digested fragments of about 256 bp on average. The genomic DNA fragments are then treated with 3' TdT to attach a poly-T tail to each fragment. A ligation containing (i) a bridging oligonucleotide with a poly-A tail, (ii) a capture sequence oligonucleotide (with one end complementary to the bridging oligonucleotide), and (iii) the T-tailed fragments is then performed, resulting in each genomic DNA fragment having the same unique capture sequence at its 3' end. Each sample of genomic DNA (*i.e.*, the test and the reference samples of nucleic acids) is coupled to a unique capture sequence prior to hybridization. Following hybridization, the genomic DNA fragments can then be labeled using a dendrimer that contains an oligonucleotide complementary to the unique capture sequence of a one sample and multiple copies of label, typically fluorescent dye molecules.

[00110] Alternatively, genomic mRNA is first reverse transcribed with unlabelled dATP, dTTP, dGTP and dCTP using a primer oligonucleotide that contains a unique capture sequence and a poly-T tail to hybridize to the poly-A tail of the mRNA molecules. The reaction is then stopped and the mRNA is degraded to yield genomic cDNAs containing the unique capture sequence. These genomic cDNAs can then be labeled using a dendrimer that contains an oligonucleotide complementary to the unique capture sequence and multiple copies of label, typically fluorescent dye molecules. Genisphere, Inc. offers a variety of dendrimers that vary in size and fluorescence intensity. The Array 900 and 350 series kits contain four-layer dendrimers. A four layer dendrimer theoretically has 324 single stranded DNA arms in the outer layer. The diameter of a four layer dendrimer is 182-190 nm and the molecular weight is 1.2×10^7 Daltons. The Array 50 series kit contains a two layer dendrimer. A two layer dendrimer theoretically has 45 single stranded DNA arms in the outer layer. The diameter of a two layer dendrimer is 70-90 nm and the molecular weight is 1.3×10^6 Daltons.

EXAMPLE 3

Comparative Genomic Hybridization

[00112] Genomic nucleic acids obtained from a test sample and a reference sample, each population containing a unique capture sequence, are combined (about 1-2 µg each) with Cot-1 DNA (80-100 µg) and precipitated with ethanol. Precipitate is collected by centrifugation and allowed to air dry for 10 minutes before re-dissolving it in a 50 µl hybridization mixture containing 50% formamide, 2 x SSC, 10% dextran sulfate, 4% SDS and 500 µg yeast tRNA, pH 7. The hybridization mixture is incubated at 70 °C for 10-15 minutes to denature the DNA and subsequently at 37 °C for 60 minutes to allow blocking of repetitive sequences. To the array is added 50 µl of slide blocking solution containing 500 µg salmon sperm DNA in 50% formamide, 2 x SSC, 10% dextran sulfate and 4% SDS, pH 7. After a 30 minute incubation at room temperature, approximately three-quarters of the blocking solution is removed, and the denatured and re-annealed hybridization mixture is added and hybridized at 37 °C for 16-72 hours. After hybridization, excess hybridization fluid is rinsed off with 0.1 M sodium phosphate, 0.1% NP40, pH 8, then the array is washed once in 50% formamide, 2 x SSC, pH 7 at 45 °C for 15 minutes, and finally with 0.1 M sodium phosphate, 0.1% NP40, pH 8 at room temperature for 15 minutes.

EXAMPLE 4

Single label CGH with Subtractive Labeling

[00113] An exemplary selective removal can be achieved by making the label associated with either the genomic nucleic acids obtained from the test sample or the genomic nucleic acids obtained from the reference sample susceptible to removal with atmospheric ozone. Certain fluorophores (*e.g.*, Cy5TM and Alexa 647) are susceptible to ozone levels as low as about 5-10 ppm for periods as short as 10-30 seconds. Following hybridization, arrays are placed in an enclosed chamber with an ozone generator to achieve at atmospheric ozone level of about 60-85 ppm for about 10-30 minutes. Selective removal of the label from one population of genomic nucleic acids may be achieved by modifying the physical nature of the labeling process, such as increasing the distance of the label from the genomic DNA to increase exposure to the atmospheric ozone.

[00114] Another exemplary selective removal can be achieved by making the label associated with either the genomic nucleic acids obtained from the test sample or the genomic

nucleic acids obtained from the reference sample susceptible to removal by cleavage with a restriction endonuclease or a homing endonuclease. In this example, reference sample genomic nucleic acids are prepared with a first unique capture sequence to which is linked a dendrimer containing an oligonucleotide complementary to this first unique capture sequence and a fluorescent label. The test sample genomic nucleic acids are prepared with a second unique capture sequence containing a stretch of nucleotides representing the recognition sequence for an endonuclease to which is linked a dendrimer containing an oligonucleotide complementary to this second unique capture sequence and the same fluorescent label as used for the first sample. Following hybridization of the test and reference genomic nucleic acids to an array containing a plurality of immobilized nucleic acid segments of interest, the fluorescence of the array is measured.

[00115] The array is then contacted with the endonuclease recognizing the sequence contained within the second unique capture sequence under conditions allowing cleavage of the dendrimeric construct from the genomic nucleic acids to selectively remove the fluorescent label from the test sample nucleic acids.

[00116] Another exemplary selective removal can be achieved by making the label associated with either the genomic nucleic acids obtained from the test sample or the genomic nucleic acids obtained from the reference sample susceptible to removal by UV irradiation. The label is incorporated using a linker that is photocleavable, such as a linker containing a 2-nitrobenzyl group (see, *e.g.*, Bai *et al.*, *Proc. Natl. Acad. Sci.* **100**:409-413, 2003). Following hybridization, arrays are placed in a chamber with water and irradiated with a UV lamp at 340 nm (light intensity of about 20 mW/cm²) for about 5-10 minutes to selectively remove the label from one population of genomic nucleic acids only (*i.e.*, the nucleic acids containing the photocleavable linker).

[00117] Thus, in these examples of selective removal, data from the array is acquired at two time points, with the same fluorophore being read. The first acquisition is after the comparative genomic hybridization (*e.g.*, before the selective removal of the label from the test sample genomic nucleic acids), in part to determine the fluorescence of the combined nucleic acid samples (F_{Total}). The second acquisition is after the selective removal of the label, in part to determine the remaining fluorescence of the reference sample genomic nucleic acids ($F_{\text{Reference}}$). The fluorescence of the test sample genomic nucleic acids (F_{Test}) is then equal to ($F_{\text{Total}} - F_{\text{Reference}}$). Thus, the same fluorophore can be used to achieve maximal

uniformity between the two genomic nucleic acid samples, and between tests performed with different samples. If the selective removal is designed to remove nucleic acid associated with the reference genomic DNA, then the second read would be F_{test} and the difference between F_{Test} and F_{Total} would be $F_{\text{Reference}}$.

[00118] As a quality control in single label CGH the two linkers for the test and reference labels are switched and comparative hybridization repeated.

EXAMPLE 5

Single label CGH with Additive Labeling

[00119] Exemplary additive labeling for single label CGH can be achieved by performing a first comparative hybridization wherein the genomic nucleic acids obtained from the reference sample comprise a first unique oligonucleotide tag and the genomic nucleic acids obtained from the test sample comprise a second unique oligonucleotide tag. Following hybridization of the test and reference genomic nucleic acids to an array containing a plurality of immobilized nucleic acid segments of interest, the array is exposed to a first dendrimeric complex containing an oligonucleotide complementary to the first unique oligonucleotide tag and a fluorescent label. This provides a selective labeling of the reference sample genomic nucleic acids.

[00120] Preferred conditions for dendrimer hybridization include use of Pronto!TM hybridization buffer (Corning, Inc.) with 50 μg of Cot I DNA and 50-100 μg of SST (shredded (sonicated) salmon testis DNA). Cot I DNA may be replaced by any other non-mammalian genomic DNA such as plant DNA, fish DNA, bacterial DNA, and non-natural DNA, e.g. dendrimeric DNA. After 30 min. hybridization, the array is washed as follows:

1. Soak slide in 2XSSC containing 0.01% SDS (pH 7.5 - 8.0) at room temperature until coverslip is loosened (<3 minutes).
2. Incubate for 5 min. with gentle agitation at 50 C in 2X SSC containing 0.01% SDS (pH 7.5 - 8.0).
3. Incubate for 5 min. with gentle agitation at room temperature in 2X SSC (pH 7.5 - 8.0).
4. Incubate for 5 min. with gentle agitation at room temperature in 0.2X SSC (pH 7.5 - 8.0).

SDS: sodium doedecyl sulfate (detergent)

1X SSC: 0.15 molar sodium chloride and 0.015 molar sodium citrate

[00121] Data from the array is then acquired, in part to determine the fluorescence of the first reference sample genomic nucleic acids ($F_{\text{Reference}}$). The array is then exposed to a second dendrimeric complex containing an oligonucleotide complementary to the second unique oligonucleotide tag and the same fluorescent label as used in the first dendrimeric complex. Data from the array is then acquired for a second time, in part to determine the fluorescence of the combined nucleic acids (F_{Total}). The fluorescence of the test sample genomic nucleic acids (F_{Test}) is then equal to ($F_{\text{Total}} - F_{\text{Reference}}$). Thus, the same fluorophore can be used to achieve maximal uniformity between the two genomic nucleic acid samples, and between tests performed with different samples. If the first dendrimeric complex binds to F_{test} , then the difference between F_{Test} and F_{Total} would be $F_{\text{Reference}}$. This method is depicted schematically in FIG. 3. A variant of this method where one labeled dendrimer is hybridized together during the hybridization with the tag labeled genomic nucleic acids is shown in FIG. 4.

[00122] As a quality control in single label CGH the unique tag sequences attached to the test and reference genomic nucleic acids are switched and comparative hybridization repeated.

[00123] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs.

[00124] The inventions illustratively described herein may suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Thus, for example, the terms “comprising”, “including,” “containing”, etc. shall be read expansively and without limitation. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed.

[00125] Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification, improvement and variation of the inventions embodied therein herein disclosed may be

resorted to by those skilled in the art, and that such modifications, improvements and variations are considered to be within the scope of this invention. The materials, methods, and examples provided here are representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention.

[00126] The invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

[00127] In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group.

[00129] Other embodiments are set forth within the following claims.

What is claimed is:

1. A method of performing comparative hybridization, said method comprising
 - (i) simultaneously hybridizing test and reference nucleic acids to a single nucleic acid array, wherein said array comprises a plurality of individual spots, wherein each spot comprises a plurality of identical nucleic acid probes and wherein said test and reference nucleic acids are labeled with the same detectable label;
 - (ii) determining a signal for the detectable label hybridized to a plurality of individual spots on the array representing the total of hybridized test and reference nucleic acid to each individual spot;
 - (iii) treating the hybridized nucleic acids to selectively remove said detectable label from one of said test or reference nucleic acids, wherein said detectable label is linked to the test and reference nucleic acids through a first and second chemical linkage, respectively, and wherein one of said first linkage or said second linkage is susceptible to the treatment step;
 - (iv) determining a signal for the detectable label hybridized to a plurality of individual spots on the array following step (iii), which represents one of said hybridized test or reference nucleic acid to each individual spot;
 - (v) determining a signal for the other of the hybridized test or reference by using the signals determined in steps (iv) and (ii); and
 - (vi) comparing the amount of test and reference nucleic acids hybridized to the plurality of individual spots.
2. The method of claim 1 wherein said detectable label is attached to the test and reference nucleic acids before hybridization.
3. The method of claim 1 wherein said detectable label is attached to the test and reference nucleic acids after hybridization.

4. The method of claim 1 wherein said detectable label is attached to the test and reference nucleic acids via a ligand-receptor interaction or by hybridization between complementary nucleotide sequences.

5. A method of performing comparative hybridization, said method comprising:

(i) hybridizing test and reference nucleic acids to a single nucleic acid array, wherein said array comprises a plurality of individual spots, wherein each spot comprises a plurality of identical nucleic acid probes;

(ii) determining a signal representing one of the test or reference nucleic acid hybridized to a plurality of individual spots on the array by detecting a first complex comprising a detectable label and a first entity, wherein said first complex selectively binds one of said test or reference nucleic acid via a first tag attached to said test or reference nucleic acid, said first entity and tag being members of a specific binding pair;

(iii) determining a signal representing the total of the test and reference nucleic acid hybridized to each individual spot on the array by contacting the hybridized array of step (ii) with a second complex comprising the detectable label of step (ii) and a second entity, wherein said second complex selectively binds the other of said test or reference nucleic acid via a second tag attached to the other of said test or reference nucleic acid detected in step (ii), said second entity and second tag being members of a specific binding pair;

(iv) using the signal from steps (ii) and (iii) to obtain the signal for the other of said hybridized test or reference nucleic acid not detected in step (ii); and

(v) comparing the amount of the test and reference nucleic acids hybridized to the plurality of individual spots on the array.

6. The method of claim 5 wherein the first complex binds to the test or reference nucleic acid detected in step (ii) before hybridization to the plurality of individual spots on the array.

7. The method of claim 5, wherein the first complex binds to the test or reference nucleic acid detected in step (ii) after hybridization to the plurality of individual spots on the array.

8. The method of any one of claims 5-7, wherein said first tag and said second tag each comprises a unique oligonucleotide capture sequence.

9. The method of claim 8, wherein said first complex and said second complex each comprises an oligonucleotide complementary to one of said unique oligonucleotide capture sequences.

10. The method of claim 8, wherein said first complex and said second complex each comprises an oligonucleotide complementary to one of said unique oligonucleotide capture sequences in a dendrimeric construct.

11. The method of any one of claims 5-7, wherein said first entity and said first tag are each attached to a member of a first specific binding pair and said second entity and second tag are each attached to a member of a second specific binding pair wherein said first and second binding pair members do not cross-react.

12. The method of any one of claims 1-11 wherein said test or reference nucleic acids represent a subset of the nucleic acid of a cell.

13. The method of any one of claims 1-11 wherein said nucleic acid array comprises a plurality of nucleic acid probes each contained within an artificial chromosome.

14. The method of claim 13, wherein said artificial chromosome is a bacterial artificial chromosome (BAC).

15. The method of claim 13, wherein said artificial chromosome is P-1 derived artificial chromosome (PAC).

16. The method of any one of claims 1-15 wherein said nucleic acid array comprises a plurality of nucleic acid probes, each nucleic acid probe of said plurality being between 1,000 and 1,000,000 nucleotides in length.

17. The method of any one of claims 1-16 wherein said nucleic acid array comprises a plurality of nucleic acid probes, each nucleic acid probe of said plurality being between 100,000 and 300,000 nucleotides in length.

18. The method of claim any one of claims 1-17, wherein said nucleic acid array comprises a plurality of nucleic acid probes that collectively comprise the nucleic acid sequence of at least one chromosome.

19. The method of claim 18, wherein said plurality of nucleic acid probes comprise segments of chromosomal sequence of 3-4 megabases in length.

20. The method of claim 18, wherein said plurality of nucleic acid probes comprise segments of chromosomal sequence of 2-3 megabases in length.

21. The method of claim 18, wherein said plurality of nucleic acid probes comprise segments of chromosomal sequence of 1-2 megabases in length.

22. The method of claim 18, wherein said plurality of nucleic acid probes collectively comprise the nucleic acid sequence of all chromosomes of a cell.

23. The method of any one of claims 1-7, wherein said nucleic acid array comprises a plurality of nucleic acid probes that collectively comprise at least 50 genomic regions of interest.

24. The method of any one of claims 1-7, wherein said nucleic acid array comprises at least 300 individual spots.

25. The method of any one of claims 1-7, wherein said nucleic acid array comprises at least 500 individual spots.

26. The method of any one of claims 1-25, wherein said detectable label is detected using optical means.

27. The method of any one of claims 1-25, wherein said detectable label is a fluorophore.

28. The method of any one of claims 1-4, wherein said first linkage and said second linkage differ in susceptibility to temperature.

29. The method of claim 28, wherein said first linkage and said second linkage both comprise oligonucleotides with different melting temperatures (T_m).

30. The method of any one of claims 1-4, wherein said first linkage and said second linkage differ in susceptibility to pH hydrolysis.

31. The method of claim 30, wherein said first linkage and said second linkage both comprise peptides.

32. The method of any one of claims 1-4, wherein said first linkage and said second linkage differ in susceptibility to radiation.

33. The method of claim 32, wherein said first linkage and said second linkage both comprise oligonucleotides comprising nucleotide stretches having different susceptibility to ultraviolet radiation.

34. The method of any one of claims 1-4, wherein said first linkage and said second linkage differ in susceptibility to a change in oxidative conditions.

35. The method of any one of claims 1-4, wherein said first linkage and said second linkage differ in susceptibility to a change in atmospheric conditions.

36. The method according to claim 35, wherein said change in atmospheric conditions comprises exposing said array to ozone.

37. The method of any one of claims 1-4, wherein either said first linkage or said second linkage is hydrolyzed upon contact with an agent causing hydrolysis.

38. The method of any one of claims 1-4, wherein either said first linkage or said second linkage is hydrolyzed upon contact with an enzyme.

39. The method of claim 38, wherein either said first linkage or said second linkage comprises an oligonucleotide containing a recognition site for a restriction endonuclease or a homing endonuclease.

40. The method of claim 39, wherein either said first linkage or said second linkage is cleaved by a rare-cutting endonuclease to selectively remove the detectable label from either nucleic acids of said test sample or nucleic acids of said reference sample upon cleavage.

41. The method of claim 39, wherein said oligonucleotide containing a recognition site for a restriction endonuclease or a homing endonuclease is in a dendrimeric construct.

42. The method of any one of claims 1-4, wherein said first linkage and said second linkage differ in susceptibility to a change in buffer conditions.

43. The method of any one of claims 1-42, wherein said test nucleic acids are comprised in a test sample that has been obtained from a patient.

44. The method of claim 43, wherein said test sample obtained from a patient is suspected of comprising cancerous cells.

45. The method of claim 43 or 44, wherein said test sample is a prenatal specimen.

46. The method of claim 43 or 44, wherein said test sample is obtained from an embryo or fetus.

47. The method of claim 43 or 44, wherein said test sample is from an individual and the method is used to detect a genetic abnormality.

48. The method of claim 43 or 44, wherein said test sample is from an individual and the method is used to diagnose a genetic abnormality.

49. The method of any one of claims 1-48, wherein said test nucleic acids and reference nucleic acids are genomic nucleic acids.

50. The method of any one of claims 1-49 wherein said test nucleic acids are cDNA prepared from mRNA.

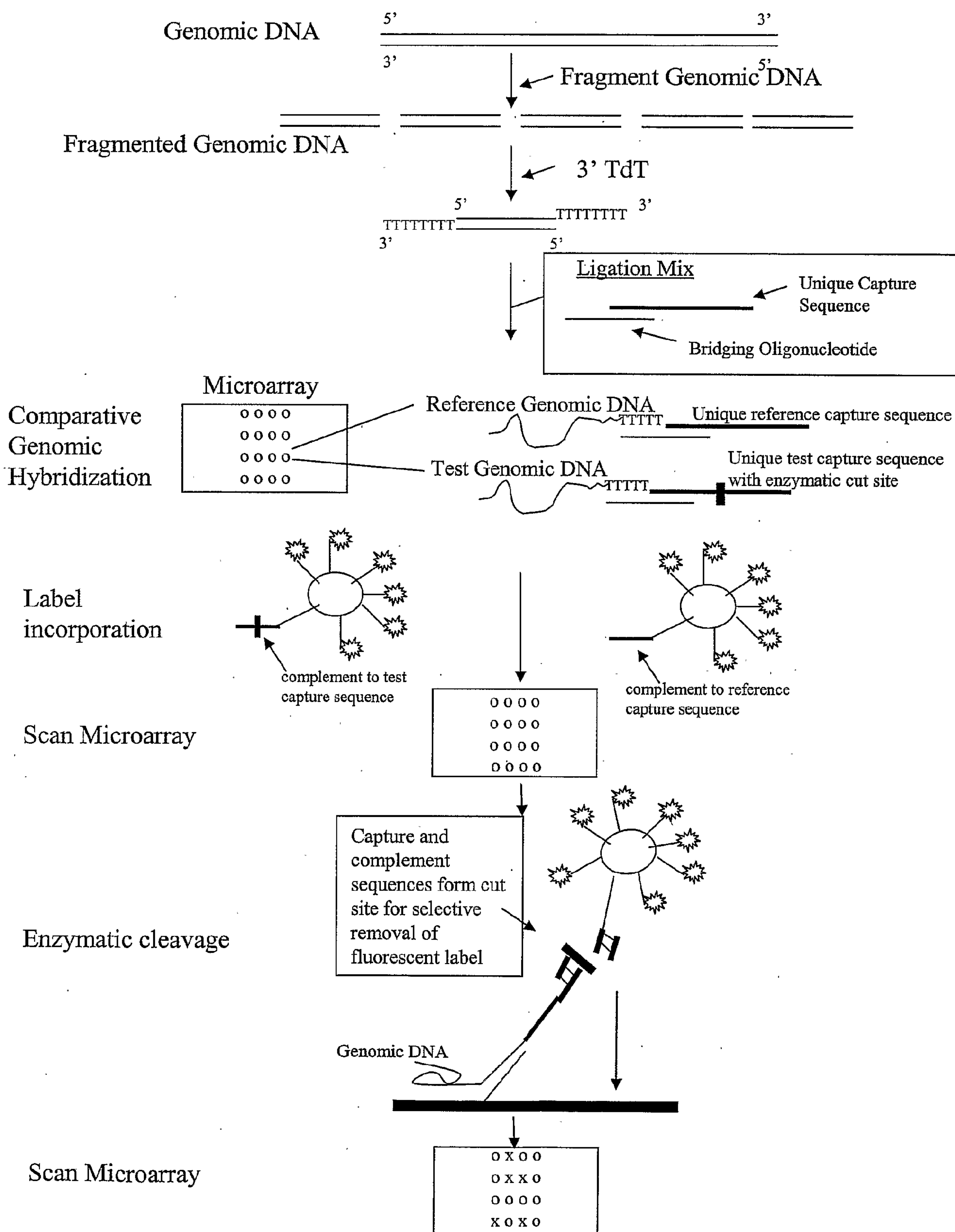
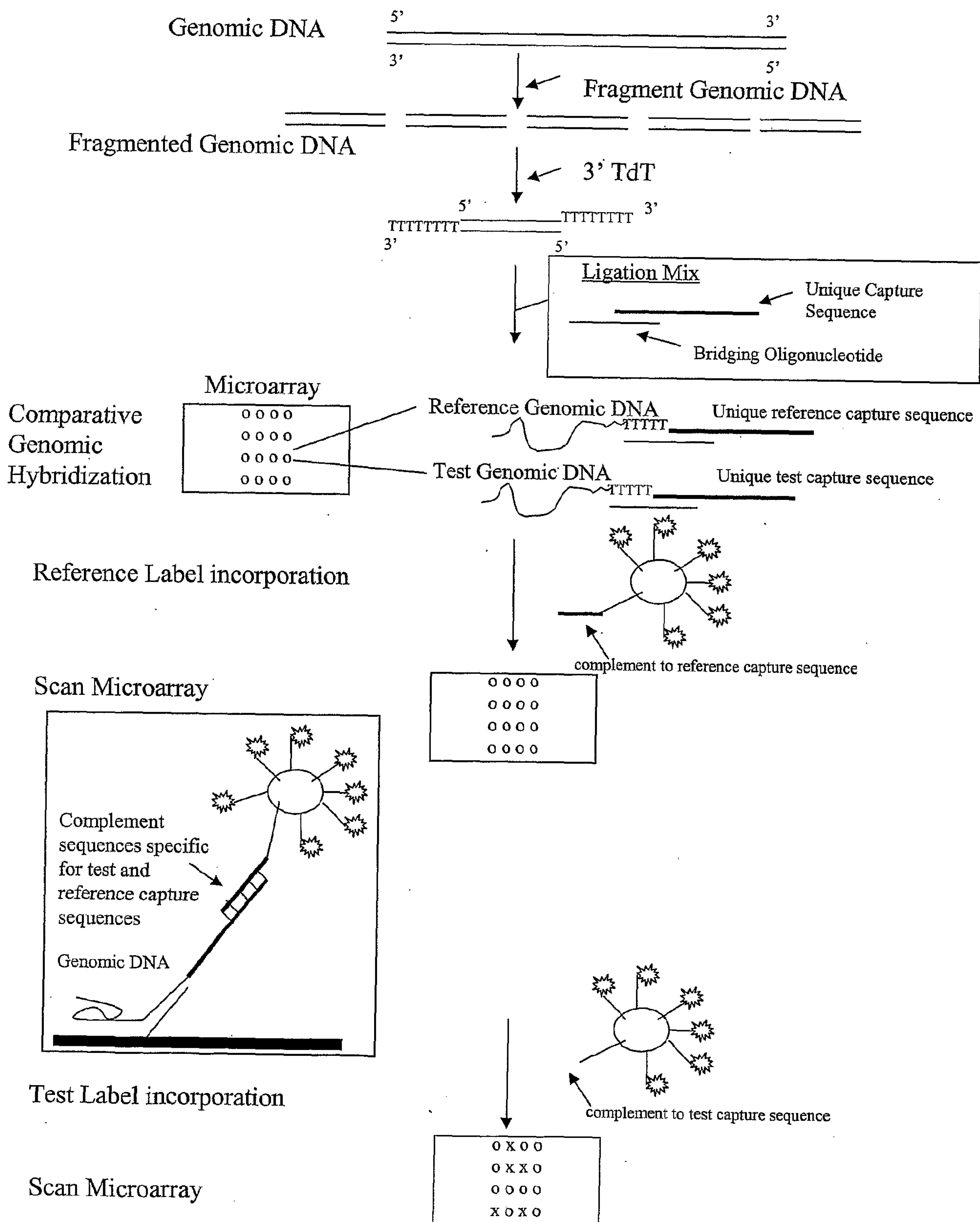
FIG. 1**1/4**

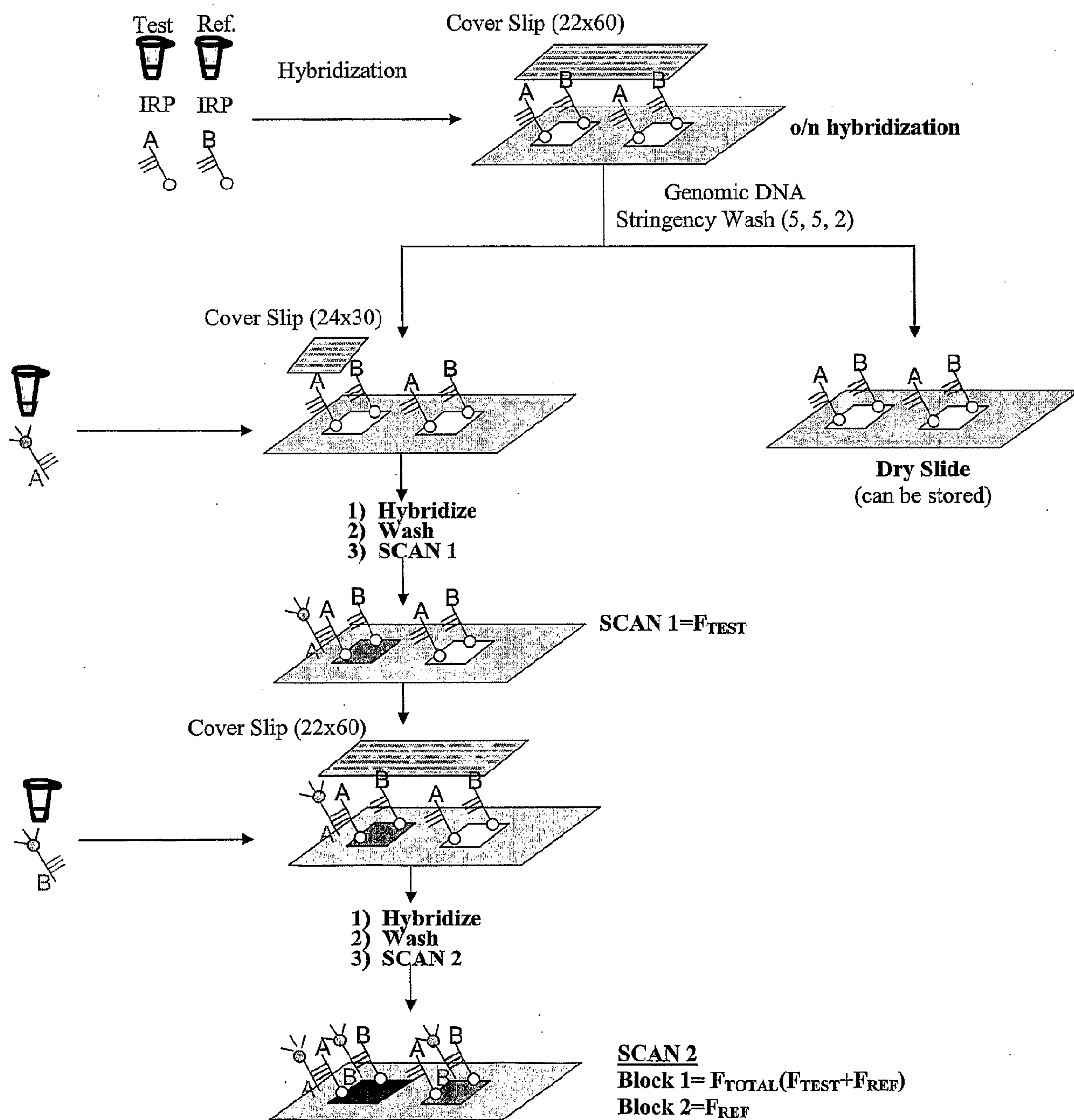
FIG. 2

2/4



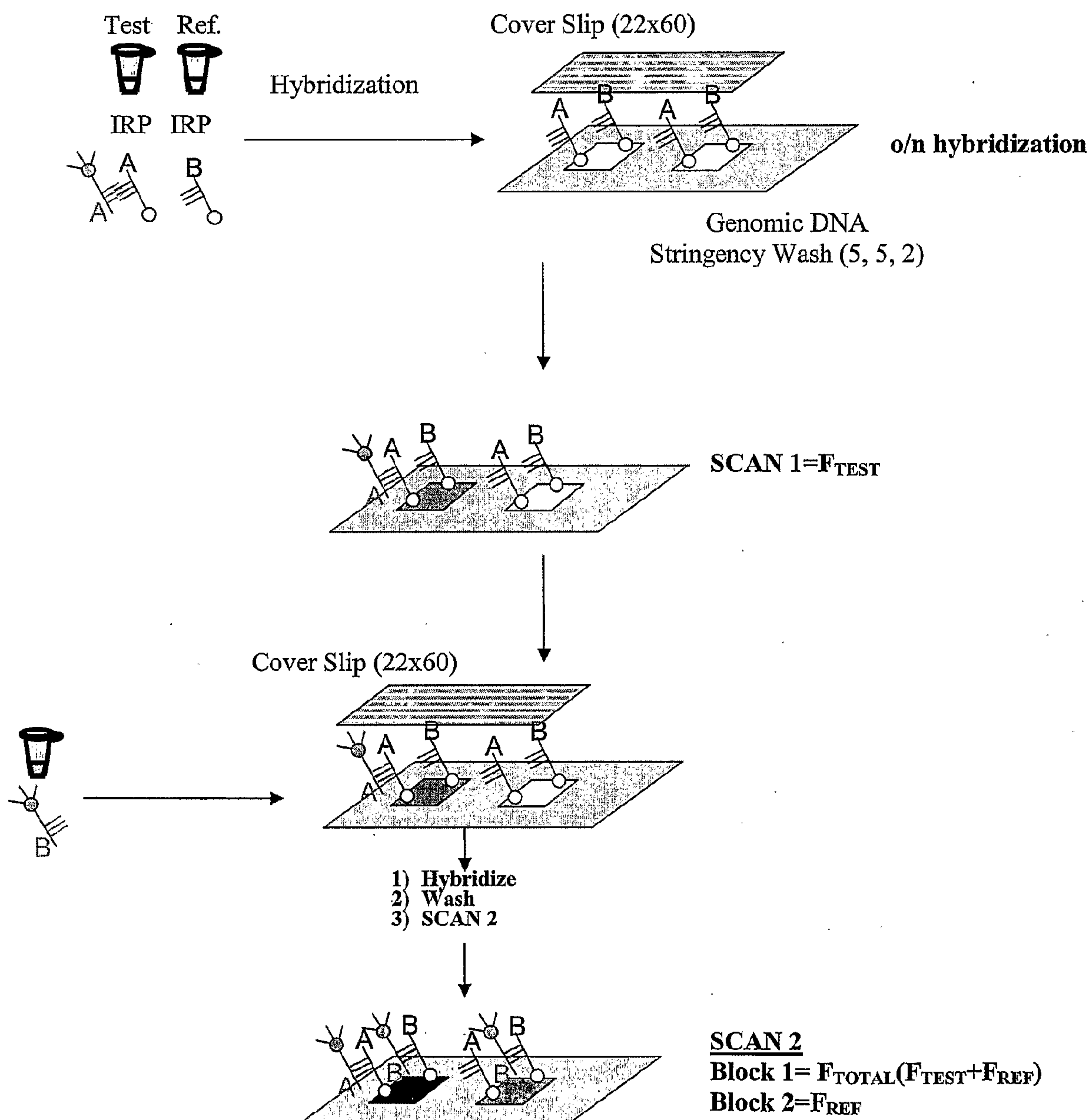
3/4

FIG. 3

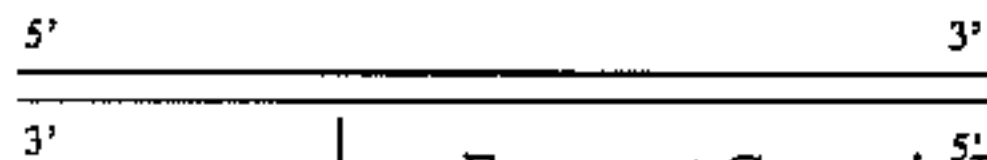


4/4

FIG. 4



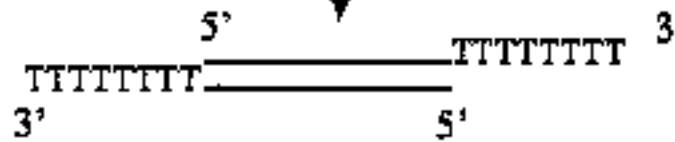
Genomic DNA



Fragment Genomic DNA

Fragmented Genomic DNA

3' TdT



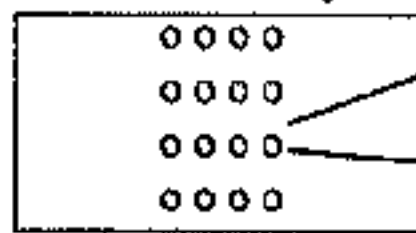
Ligation Mix

Unique Capture Sequence

Bridging Oligonucleotide

Comparative Genomic Hybridization

Microarray



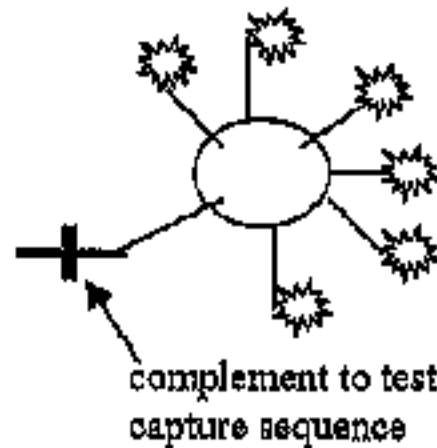
Reference Genomic DNA

Unique reference capture sequence

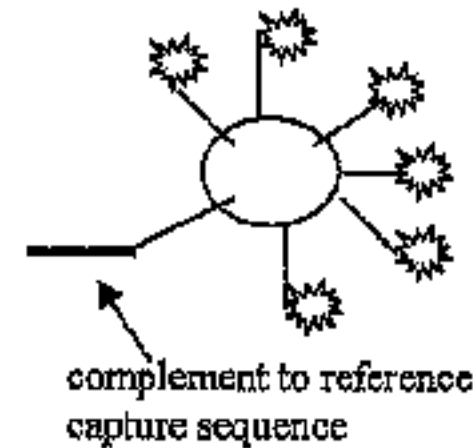
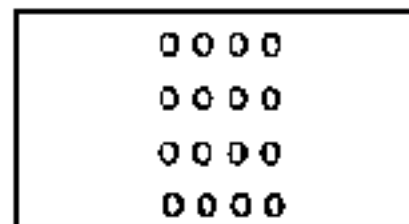
Test Genomic DNA

Unique test capture sequence with enzymatic cut site

Label incorporation



Scan Microarray



Enzymatic cleavage

Capture and complement sequences form cut site for selective removal of fluorescent label

Genomic DNA



Scan Microarray

