DETECTION OF TARGET NUCLEIC ACIDS USING HYBRIDIZATION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present international PCT application claims the benefit of US Patent Application No. 14/450,144, filed August 1, 2014 and US Patent Application No. 14/453,396, filed August 6, 2014.

FIELD OF THE INVENTION

[0002] This invention relates to detection of target genomic regions from samples.

BACKGROUND OF THE INVENTION

[0003] In the following discussion certain articles and methods will be described for background and introductory purposes. Nothing contained herein is to be construed as an "admission" of prior art. Applicant expressly reserves the right to demonstrate, where appropriate, that the articles and methods referenced herein do not constitute prior art under the applicable statutory provisions.

[0004] Genetic abnormalities account for a wide number of pathologies, including pathologies caused by chromosomal aneuploidy (*e.g.*, Down syndrome), germline mutations in specific genes (*e.g.*, sickle cell anemia), and pathologies caused by somatic mutations (*e.g.*, cancer). Diagnostic methods for determining genetic anomalies have become standard techniques for identifying specific diseases and disorders, as well as providing valuable information on disease source and treatment options.

[0005] Copy number variations (CNVs) are alterations of genomic DNA that correspond to specific regions of the genome—including entire chromosomes—that have been deleted or duplicated. CNVs can be caused by genomic rearrangements such as deletions, duplications, inversions, and translocations. CNVs have been associated with various forms of cancer (Cappuzzo F, Hirsch, et al. (2005) *J Natl Cancer Inst.,* 97(9):643-55), neurological disorders,

including autism (Sebat, J., et al. (2007) *Science* 316(5823):445-9), and schizophrenia (St. Clair, D., (2008). *Schizophr Bull* 35(1):9-12).

[0006] Therefore, there is a need for methods of screening for copy number variations that employs an efficient, reproducible assay and detection system.

SUMMARY OF THE INVENTION

[0007] This Summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This Summary is not intended to identify key or essential features of the claimed subject matter, nor is it intended to be used to limit the scope of the claimed subject matter. Other features, details, utilities, and advantages of the claimed subject matter will be apparent from the following written Detailed Description including those aspects illustrated in the accompanying drawings and defined in the appended claims.

[0008] The present invention provides methods for detecting genetic characteristics in a sample, including copy number variations (CNVs), insertions, deletions, translocations, polymorphisms and mutations. The invention employs the technique of interrogating loci from two or more target genomic regions using at least two fixed sequence oligonucleotides for each interrogated locus, and joining the fixed sequence oligonucleotides either directly or indirectly via ligation. The ligation products from different loci in a selected genomic regions comprise nucleic acid capture regions designed to include a region complementary to one or more capture probes on a solid support. The capture region comprises one or more detectable labels that identify a ligation product as originating from a specific target genomic region. Identification of ligation products from different target genomic regions is achieved by binding of the capture regions of the ligation products to complementary capture probes on the solid support.

[0009] In a specific embodiment, the invention provides an assay method for providing a statistical likelihood of a fetal aneuploidy comprising providing a maternal sample comprising

maternal and fetal cell free DNA, interrogating one or more loci from a first target genomic region using sequence-specific oligonucleotides that comprise a capture region, interrogating one or more loci from a second target genomic region using sequence-specific oligonucleotides that comprise a capture region, detecting the isolated selected loci from the first and second target genomic regions via hybridization to an array, quantifying total counts of the isolated loci to determine a relative frequency of first and second target genomic regions, interrogating selected polymorphic loci from at least one target genomic region different from the first and second target genomic regions using sequence-specific oligonucleotides, detecting isolated selected polymorphic loci, quantifying total counts of the isolated selected polymorphic loci to calculate a percentage of the fetal cell free DNA in the maternal sample, calculating a statistical likelihood a fetal aneuploidy in the maternal sample, wherein the relative frequency of loci from the first target genomic region, the relative frequency of loci from the second target genomic region, and quantified counts from the isolated selected polymorphic loci to provide a statistical likelihood of the presence of a fetal aneuploidy.

[00010] In other specific embodiments, the invention provides an assay method determining the presence or absence of a fetal aneuploidy comprising providing a maternal sample comprising maternal and fetal cell free DNA, interrogating one or more loci from a first target genomic region using sequence-specific oligonucleotides that comprise a capture region, interrogating one or more loci from a second target genomic region using sequence-specific oligonucleotides that comprise a capture region, detecting the loci from the first and second target genomic regions via hybridization to an array, quantifying total counts of the loci to determine a relative frequency of first and second target genomic regions, and determining the presence or absence of a fetal aneuploidy in the maternal sample based on a deviation from expected counts of the isolated loci using the relative frequency of loci from the first target genomic region and the relative frequency of loci from the second target genomic region. In certain aspects, the deviation from expected counts is determined using a threshold level determined from a

representative population of samples, and preferably a representative population comprising samples from patients of similar maternal age and/or gestational age.

[00011] In specific aspects, the interrogation of the loci from the first and second genomic regions uses hybridization followed by ligation. In more specific aspects, an amplification step is performed after the hybridization and ligation steps. In other specific aspects, the amplification is universal amplification using the polymerase chain reaction.

[00012] In preferred aspects, the ligation products from two or more different genomic regions are identified using a single solid support with capture probes; *e.g.,* an array comprising capture probes complementary to multiple capture regions indicative of the different target genomic regions. Upon introducing a pool of ligation products originating from two or more different genomic regions to the array, ligation products having the same capture region will competitively hybridize to complementary capture probes on the array, and the relative frequency of ligation products from each genomic region can be estimated based on the amount of detected label bound to the capture probes. In this manner, the relative frequencies of the target genomic regions themselves may be determined. The relative frequencies of each target genomic region may be determined by identifying the binding of capture regions on the ligation products corresponding to each selected locus from each target genomic region to specific, known locations on the array, or by estimating total fluorescence from the array following binding of the ligation products originating from the target genomic regions.

[00013] In certain preferred embodiments, the capture regions and capture probes do not reflect the specific target genomic region nucleotide sequence, and are instead "engineered" sequences that serve as surrogates to identify specific target genomic regions; thus, the nucleotide sequence of the ligation product corresponding to the target genomic region does not need to be determined directly. The use of the capture regions on the ligation products allows the binding of the ligation products to the capture probes on the array to indicate the larger target genomic region from which the ligation product originates without the need to sequence

the portion of the ligation product corresponding to the actual nucleotide sequence of the target genomic region. Because the capture regions and capture probes are engineered sequences, they can be thought of as "universal" sequences; that is, these capture regions and capture probes can be used in conjunction with any number of different assays, the only difference being the target sequence(s) associated with the capture region(s).

[00014] In one embodiment, the arrays of the present invention comprise capture probes that all have substantially the same sequence. In another embodiment, the arrays used comprise two to several different features with capture probes having substantially the same sequence. These arrays are in contrast to arrays known in the art that identify individual sequences by complementarity to individual features with each feature comprising a nucleic acid sequence different from the other features. The use of a single or a limited number of complementary capture probe sequences in the individual features on an array can simplify the biochemistry needed to create the array and reduce potential spurious differences in detection frequency resulting, *e.g.,* from differences in binding affinity between the capture regions on the ligation products and the capture probes.

[00015] In specific embodiments, the arrays comprise two or more different capture probes used to detect individual ligation products from two or more different target genomic regions, two or more different loci from a single target genomic region, or two or more different alleles from a selected locus. That is, capture probes of different sequence hybridize to capture regions on ligation products that correspond to different target genomic regions, different loci from a single target genomic region, or different alleles from a selected locus. The capture regions on the ligation products are associated with labels indicative of the target genomic region or selected locus, or indicative of the alleles of a polymorphism, from which the ligation product originated.

[00016] Thus, it is a preferred embodiment of the invention that the ligation products are identified using a capture probe that is complementary to a capture region introduced in or to the ligation product, but that does not identify the target genomic region to which the ligation

product corresponds solely by hybridization to a feature complementary to the target genomic region. In some embodiments, the capture probe is in part complementary to a target genomic region and in part complementary to a capture region in a ligation product. In a preferred embodiment, the capture region used to identify a ligation product which corresponds to a target genomic region is not complementary to any portion of the target genomic region.

[00017] Preferably the capture region is introduced as part of one of the fixed sequence oligonucleotides prior to ligation, although the capture region may be attached (*e.g.,* via ligation of an adaptor) to one or both ends of the ligation product of the fixed sequence oligonucleotides following the ligation procedure.

[00018] It is another feature of the hybridization assay format of the invention that quantification of target genomic regions when using capture regions can be achieved by quantifying labels that are associated with the ligation products from the loci within the target genomic region without actually determining the sequence of the ligation products corresponding to the target genomic regions. In this manner, the frequency of ligation products from a target genomic region (and thus the frequency of the target genomic regions themselves) can be estimated without the need to detect the actual nucleotide sequence of the loci from that target genomic region.

[00019] In many preferred embodiments, quantification of the labels bound to the capture probes on the array is the only readout necessary to estimate the levels or amounts of ligation products produced from each target genomic region, which in turn can be used to estimate the frequency of the target genomic regions.

[00020] It is an advantage of the methods of the invention that the ligation products can be associated with multiple different detectable labels and/or capture regions. The use of different labels and/or capture regions in different experiments can mitigate any frequency bias from the use of a particular detectable label, capture region or capture probe.

[00021] In certain embodiments, the present invention provides methods for detecting frequencies of first and second target genomic regions in a sample comprising: introducing a first set of first and second fixed sequence oligonucleotides to a sample under conditions that allow the first and second fixed sequence oligonucleotides to hybridize specifically to complementary regions in loci from a first target genomic region, wherein at least one of the first or second fixed sequence oligonucleotide of the first set comprises a capture region and a first label; introducing a second set of first and second fixed sequence oligonucleotides to a sample under conditions that allow the first and second fixed sequence oligonucleotides to hybridize specifically to complementary regions in loci from a second target genomic region, wherein at least one of the first or second fixed sequence oligonucleotide of the second set comprises a capture region and a second label; ligating the hybridized fixed sequence oligonucleotides to create ligation products complementary to the loci; introducing the ligation products to an array comprising capture probes under conditions that allow the capture probes on the array to hybridize specifically to the capture regions of the ligation products; detecting the first and second labels; and quantifying a relative frequency of the first and second labels to quantify the relative frequency of the first and second target genomic regions. In some aspects, the capture regions of the first and second sets of ligation products are different. In other aspects the capture regions of the first and second sets of ligation products are the same.

[00022] In a preferred aspect, the first and second capture regions of the first and second sets of oligonucleotides are the same. In other aspects, a small number of capture regions are used in the fixed sequence oligonucleotides of both the first and the second set. In other aspects, the first capture regions corresponding to the first target genomic region are different from the second capture regions corresponding to the second genomic region.

[00023] In other preferred embodiments, in addition to interrogating first and second target genomic regions, polymorphic sequences comprising SNPs from two or more selected polymorphic loci also are interrogated. Selected polymorphic loci are interrogated using a third

set of fixed sequence oligonucleotides to determine allele frequencies. The allele frequencies are used, *e.g.,* to calculate the percent of fetal nucleic acids present in a maternal serum sample.

[00024] In certain embodiments, the first and second fixed sequence oligonucleotides do not hybridize adjacently to the loci in the target genomic regions, and instead have an intervening region or "gap" between the fixed sequence oligonucleotides of a set hybridized to a locus. This intervening region may be filled, *e.g.,* using a polymerase and dNTPs to extend the end of one fixed sequence oligonucleotide so that the end becomes adjacent to the end of the other hybridized oligonucleotide of the set. In another embodiment, the intervening region may be filled using one or more "gap-filling" or "bridging" oligonucleotides that bind between and adjacent to the fixed sequence oligonucleotides of a set. In the latter case, preferably the ligation step will ligate all of the oligonucleotides into a single, contiguous ligation product comprising a single capture region which can then be detected on an array. In yet another embodiment, a combination of bridging oligonucleotides and dNTPs and polymerase can be used to fill the intervening space between the fixed sequence oligonucleotides.

[00025] In the above-described embodiments, a set of fixed sequence nucleic acids is used which comprises two separate fixed sequence oligonucleotides designed to hybridize to two separate regions in each selected locus (either adjacently or non-adjacently). In some embodiments, however, a set of fixed sequence oligonucleotides can comprise a single probe with regions at either end complementary to a selected locus. Upon hybridization of this single probe to a locus, the probe forms a circular structure that may or may not be adjacently hybridized on the locus. . Such "precircle" probes can also hybridize with a gap between the ends of the probe, which gap may be filled by the hybridization of one or more bridging oligonucleotides, by extension of one end of the probe using polymerase and dNTPs, or a combination thereof.

[00026] In certain aspects, the detectable labels are directly associated with (*i.e.,* covalently or non-covalently bound to a capture region that is in one of the fixed sequence oligonucleotides of each set. In another embodiment, the ligation products are amplified following ligation, *e.g.,* in a universal amplification, and the detectable label is associated with a capture region contained within a primer used for the amplification. In other specific aspects, the isolated loci from the first and second target genomic regions and the ligation products from the selected polymorphic loci are amplified in a single vessel. In other aspects, the detectable labels are covalently or non-covalently bound to an oligonucleotide that hybridizes to a complementary sequence on the ligation products, amplicons or cleavage products thereof. Such labeled oligonucleotides may be hybridized to the ligation products prior to or after introduction of the ligation products to the capture probe array.

[00027] In other embodiments of the invention, the capture regions on the ligation products corresponding to different target genomic regions comprise different sequences, and the comparative frequency of at least a first and a second target genomic region are determined based on the use of different detectable labels associated with the different capture regions.

[00028] In certain aspects, copy number variants are is detected by an alteration of an expected ratio of bound detectable label from the bound ligation products from the target genomic regions in the sample. In certain specific aspects copy number variants are detected by an increased or decreased level of hybridization of a first set of ligation products from a first selected locus as compared to a second set of ligation products from a second selected locus.

[00029] The relative frequency of loci in a sample can be used to determine not only copy number variation for a small target genomic region, but also in conjunction with and/or in comparison to other loci, the relative frequency of loci may be used to determine the copy number variation of larger target genomic regions, including partial or whole chromosomes.

[00030] In another general aspect of the invention, a method for detecting frequencies of first and second target genomic regions in a sample is provided comprising introducing a first set of

first and second fixed sequence oligonucleotides to a sample under conditions that allow the first and second fixed sequence oligonucleotides to hybridize specifically to complementary regions in loci in the first target genomic region to create first hybridized fixed sequence oligonucleotides; introducing a second set of first and second fixed sequence oligonucleotides to the sample under conditions that allow the first and second fixed sequence oligonucleotides to hybridize specifically to complementary regions to create second hybridized fixed sequence oligonucleotides; and ligating the hybridized fixed sequence oligonucleotides of each set to create ligation products complementary to loci in the first and second target genomic regions. At least one fixed sequence oligonucleotide of each set comprises a capture region comprising a sequence complementary to capture probes on an array and a binding region for a detectable label. The ligation products are introduced to the array comprising capture probes complementary to the capture regions of the ligation products under conditions that allow the capture probes to specifically hybridize to the capture regions of the ligation products. A first detectable label is introduced to the array under conditions that allow the detectable label to specifically hybridize to the binding region of the capture region on the ligation products from the loci from the first target genomic region, and a second detectable label is introduced to the array under conditions that allow the detectable label to specifically hybridize to the binding regions of the capture region on the ligation products from the loci from the second target genomic region. The first and second detectable labels are detected and quantified to provide a relative frequency of the first and second target genomic regions in the sample. As discussed, above, in relation to one embodiment, in addition to interrogating first and second target genomic regions, aspects of this embodiment interrogate polymorphic sequences comprising SNPs from two or more selected polymorphic loci. Selected polymorphic loci are interrogated using a third set of fixed sequence oligonucleotides to determine allele frequencies. The allele frequencies are used, *e.g.,* to calculate the percent of fetal nucleic acids present in a maternal serum sample. Also as discussed, above, in some aspects of this embodiment, “gap-filling” or “bridging”

oligonucleotides may be employed in addition to the two fixed sequence oligonucleotides, and in some aspects of this embodiment, the fixed sequence oligonucleotides or the bridging oligonucleotides are allele-specific, as described in detail *infra.*

[00031] In specific aspects, the assay of the invention provides identifying low frequency alleles from the isolated selected polymorphic loci where the maternal DNA is homozygous and the non-maternal DNA is heterozygous, computing a sum of low frequency alleles from the isolated selected polymorphic loci, and calculating a statistical likelihood of a fetal aneuploidy in the maternal sample using the sum of the low frequency alleles from the isolated selected polymorphic loci to calculate statistically significant differences in target genomic region frequencies for the first and second target genomic regions, and wherein a statistically significant difference in chromosomal frequency provides a statistical likelihood of the presence of a fetal aneuploidy.

[00032] In yet another general aspect of the invention, a method is provided for detecting frequencies of first and second target genomic regions in a sample, the method comprising providing a sample; introducing a first set of first and second fixed sequence oligonucleotides to the sample under conditions that allow the fixed sequence oligonucleotides to hybridize specifically to complementary regions in first loci of the first target genomic region to create first hybridized fixed sequence oligonucleotides; and introducing a second set of first and second fixed sequence oligonucleotides to the sample under conditions that allow the fixed sequence oligonucleotides to hybridize specifically to complementary regions in second loci of the second target genomic region to create second hybridized fixed sequence oligonucleotides. At least one of the fixed sequence oligonucleotides of the first set of fixed sequence oligonucleotides comprises a capture region and a label binding region complementary to a first detectable label, and at least one of the fixed sequence oligonucleotides of the second set of fixed sequence oligonucleotides comprises substantially the same capture region as the first set and a label binding region complementary to a second detectable label. The fixed sequence

oligonucleotides of the first and second set are ligated to create first and second ligation products comprising regions complementary to the first and second loci, respectively, and labeled with a first and second detectable label. The first and second ligation products are introduced to an array comprising capture probes under conditions that allow the capture probes to specifically hybridize to the capture regions of the ligation products. The first and second labels are detected and quantified to determine a relative frequency of the first and second labels, thereby quantifying a relative frequency of the first and second target genomic regions in the sample.

[00033] The detectable labels may be introduced to the ligation products prior to the introduction of the ligation products to the array. Alternatively, the detectable labels may be introduced to the array following hybridization of the ligation products.

[00034] Again, in certain embodiments, the methods further employ the extension of at least one fixed sequence oligonucleotide hybridized to a sequence of interest. That is, in some embodiments the fixed sequence oligonucleotides that hybridize to one or more loci may not hybridize adjacently, leaving a "gap" or "intervening" region. This intervening region may be filled, *e.g.,* using a polymerase and dNTPs to extend the end of one fixed sequence oligonucleotide so that the end is adjacent to the end of the other hybridized fixed sequence oligonucleotide of the set. In another embodiment, the intervening region may be filled using one or more "gap-filling" or "bridging" oligonucleotides that bind between and adjacent to the fixed sequence oligonucleotides of a set. In the latter case, preferably the ligation step will ligate all of the oligonucleotides into a single, contiguous ligation product comprising a single capture region which can then be detected on an array. Also, a combination of bridging or gap-filling oligonucleotides and dNTPs and polymerase can be used to fill the gap. Additionally in some embodiments, pre-circular, padlock or molecular inversion probes may be used in lieu of two fixed sequence oligonucleotides in a set.

[00035] When gap-filling or bridging oligonucleotides are used, the bridging oligonucleotides typically are short, preferably between 2-30 nucleotides and more preferably between 3-28 nucleotides in length. In one aspect, the bridging oligonucleotides can be designed to provide degeneracy at multiple or all positions, *e.g.*, the bridging oligonucleotides may be full or partial randomers with various sequence variations to ensure detection of the loci even if a locus contains a polymorphic nucleotide at one or more positions. The degeneracy of the bridging oligonucleotide can be designed based on the predicted polymorphisms that may be present in the loci. Alternatively, in another aspect the pool of bridging oligonucleotides used in a reaction can provide limited degeneracy targeting specifically one or more positions based on predicted polymorphisms that may be present in the regions of the loci. In yet another aspect, the pool of bridging oligonucleotides used in a reaction can provide degeneracy for each internal position, with the nucleotides adjacent to the sites of ligation with the fixed sequence oligonucleotides remaining fixed. It is an advantage that using degenerate bridging oligonucleotides obviates the need to predetermine the maternal and fetal polymorphic content for a selected locus prior to employing the detection methods of the present invention.

[00036] In another aspect, the bridging oligonucleotide is longer than 10 nucleotides in length and is preferably 18-30 nucleotides in length. In a preferred aspect, there is a single bridging oligonucleotide complementary to each selected locus designed to hybridize between the regions of the selected locus complementary to the first and second fixed sequence oligonucleotides. In another aspect, two or more bridging oligonucleotides are designed to hybridize between the fixed sequence oligonucleotides at each selected locus, and preferably the bridging oligonucleotides hybridize adjacently to the first and second fixed sequence oligonucleotides.

[00037] In the situation where there are two bridging oligonucleotides, three ligation events occur per selected locus: ligation between the first fixed oligonucleotide and the first bridging oligonucleotide, ligation between the first and second bridging oligonucleotides, and ligation

between the second bridging oligonucleotide and the second fixed sequence oligonucleotide. In another aspect, there may be gaps between the bridging oligonucleotides and/or between the bridging oligonucleotides and the fixed sequence oligonucleotides. These gaps can be filled by extension—*e.g.,* by use of polymerase and dNTPs—prior to ligation.

[00038] In one aspect of the invention, the first and second fixed sequence oligonucleotides are introduced to the sample and specifically hybridized to the complementary portions of the loci prior to introducing the bridging oligonucleotides to the sample. In another aspect, the bridging oligonucleotides are introduced to the sample at the same time the first and second sets of fixed sequence oligonucleotides are introduced to the sample.

[00039] In another general aspect of the invention, a method for determining a presence or absence of an aneuploidy in a mixed sample is provided, the method comprising providing a mixed sample; introducing a first set of first and second fixed sequence oligonucleotides to the mixed sample under conditions that allow the fixed sequence oligonucleotides to hybridize specifically to first loci on a first chromosome to create first hybridized fixed sequence oligonucleotides, where the first set of fixed sequence oligonucleotides comprises a first capture region and a first label binding region; introducing a second set of first and second fixed sequence oligonucleotides to the mixed sample under conditions that allow the fixed sequence oligonucleotides to hybridize specifically to second loci on a second chromosome to create second hybridized fixed sequence oligonucleotides, where the second set of fixed sequence oligonucleotides comprises a second capture region and a second label binding region; ligating the hybridized oligonucleotides to create ligation products complementary to the loci; introducing the ligation products to an array comprising capture probes under conditions that allow the capture probes to hybridize specifically to the first and second capture regions of the ligation products; introducing a first labeled oligonucleotide to the array under conditions that allow a target recognition region of the first labeled oligonucleotide to hybridize specifically to the first label binding region; introducing a second labeled oligonucleotide to the array under conditions

that allow a target recognition region of the second labeled oligonucleotide to hybridize specifically to the second label binding region; detecting first and second labels; quantifying relative frequencies of the first and second labels, thereby quantifying a relative frequency of the first and second chromosome (or genomic region) in the mixed sample, wherein a statistically significant difference in the relative frequencies of the labels on the array is indicative of the presence or absence of a chromosomal aneuploidy in the mixed sample.

[00040] In alternative embodiments, a "threshold" level can be used to determine the presence or absence of a fetal aneuploidy based on the observed deviation of the relative frequency of the first and second chromosome in the mixed sample. This threshold may be determined, *e.g.,* using techniques such as those disclosed in U.S. App. 2012/0149583, 2013/0324420, 2013/0029852 and U.S. Pat. No. 8,532,936. In certain aspects, the deviation from expected counts is determined using a threshold level determined from a representative population of samples, and preferably a representative population comprising samples from patients of similar characteristics, such as prior risk profile, maternal age and/or gestational age.

[00041] In preferred aspects of the invention, the sample DNA is bound to a solid support, either before, during or after the addition of the fixed sequence oligonucleotides. In preferred aspects of the invention, the assays employ steps to remove unhybridized oligonucleotides prior to creation of ligation products, *e.g.,* by washing or by exonuclease digestion. In other preferred aspects, the ligation products are isolated following ligation but prior to further processing and/or introduction to the array for detection. In other preferred embodiments, the ligation products are amplified, preferably using universal primers, to form amplicons. In other preferred embodiments, the amplicons are subsequently cleaved to form cleaved amplicons before hybridization to an array. In embodiments involving cleavage of the ligation products, the cleaved region comprising the capture regions is preferably separated from the remainder of the cleavage products prior to introduction of the capture region portion to the array.

[00042] In certain aspects, the sample DNA, ligations products and/or the amplification products are isolated using conventional techniques in the art. For example, the hybridization complexes (*e.g.,* the fixed sequence oligonucleotides bound to the target loci), ligations products and/or the amplification products may be isolated by attachment to a solid substrate followed by a separation step, *e.g.* washing or nuclease digestion. In specific examples, they may be isolated using attachment to magnetic beads. In other specific examples, they may be isolated using attachment to a substrate with a binding partner, *e.g.* the oligonucleotide is biotinylated and the substrate comprises avidin or streptavidin. In aspects in which precircle probes are used, the hybridization complexes and/or ligation products may be isolated by nuclease destruction of non-circularlized probes.

[00043] In some aspects of this embodiment, the first and second capture regions have the same nucleotide sequence. In other aspects of this embodiment, the first and second capture regions have different nucleotide sequences. Also as discussed, above, in relation to other embodiments, in addition to interrogating first and second target genomic regions, aspects of this embodiment interrogate polymorphic sequences comprising SNPs from two or more selected polymorphic loci. Selected polymorphic loci are interrogated, *e.g.,* using a third set of fixed sequence oligonucleotides to determine allele frequencies. The allele frequencies are used, *e.g.,* to calculate the percent of fetal nucleic acids present in a maternal serum sample.

[00044] As discussed above, in some aspects of this embodiment, extension ligation and/or "bridging" oligonucleotides may be employed in addition to the two fixed sequence oligonucleotides. Accordingly, the invention provides a method for determining a likelihood of a fetal aneuploidy comprising the steps of providing a maternal sample comprising maternal and fetal cell free DNA, introducing first sets of two fixed sequence oligonucleotides complementary to loci in a first target genomic region in the maternal sample under conditions that allow a complementary region of each fixed sequence oligonucleotide to specifically hybridize to the loci, wherein at least one of the two fixed sequence oligonucleotides of each set comprises a

universal primer site and a capture region, introducing second sets of two fixed sequence oligonucleotides complementary to loci in a second target genomic region in the maternal sample under conditions that allow a complementary region of each fixed sequence oligonucleotide to specifically hybridize to the loci, wherein at least one of the two fixed sequence oligonucleotides of each set comprises a universal primer site and a capture region, introducing third sets of two fixed sequence oligonucleotides complementary to a set of polymorphic loci in a target genomic region that is different from the first target genomic region in the maternal sample under conditions that allow a complementary region of each fixed sequence oligonucleotide to specifically hybridize to selected polymorphic loci, wherein at least one of the two fixed sequence oligonucleotides of each set comprises a universal primer site and a capture region, introducing bridging oligonucleotides to the maternal sample under conditions that allow the bridging oligonucleotides to specifically hybridize to complementary regions in the loci between the fixed sequence oligonucleotides, ligating the hybridized first and second fixed sequence oligonucleotides and the bridging oligonucleotides to create ligation products complementary to the loci, isolating the ligation products, amplifying the isolated ligation products using the universal primer sites, applying the amplified ligation products to an array, wherein the array comprises capture probes complementary to the capture regions on the ligation products, quantifying a relative frequency of each allele from the selected polymorphic loci to determine a percent fetal cell-free DNA in the sample, quantifying a relative frequency of loci from the first target genomic region and a relative frequency of loci from the second target genomic region, and computing a likelihood of the presence or absence of a fetal aneuploidy using the relative frequency of the loci from the first and second target genomic regions and the percent fetal cell-free DNA to determine the likelihood of the presence or absence of a fetal aneuploidy.

[00045] The invention also provides a method for determining a likelihood of a fetal aneuploidy comprising the steps of providing a maternal sample comprising maternal and fetal cell free

DNA, introducing first sets of two fixed sequence oligonucleotides complementary to loci in a first target genomic region in the maternal sample under conditions that allow a complementary region of each fixed sequence oligonucleotide to specifically hybridize to the loci, wherein at least one of the two fixed sequence oligonucleotides of each set comprises a universal primer site and a capture region, introducing second sets of two fixed sequence oligonucleotides complementary to loci in a second target genomic region in the maternal sample under conditions that allow a complementary region of each fixed sequence oligonucleotide to specifically hybridize to the loci, wherein at least one of the two fixed sequence oligonucleotides of each set comprises a universal primer site and a capture region, introducing third sets of two fixed sequence oligonucleotides complementary to a set of polymorphic loci in a target genomic region that is different from the first target genomic region in the maternal sample under conditions that allow a complementary region of each fixed sequence oligonucleotide to specifically hybridize to selected polymorphic loci, wherein at least one of the two fixed sequence oligonucleotides of each set comprises a universal primer site and a capture region, extending at least one of the hybridized fixed sequence oligonucleotides using dNTPs and a polymerase to create adjacently hybridized oligonucleotides, ligating the adjacently hybridized oligonucleotides to create ligation products complementary to the loci, isolating the ligation products, amplifying the isolated ligation products using the universal primer sites, applying the amplified ligation products to an array, wherein the array comprises capture probes complementary to the capture regions on the ligation products, quantifying a relative frequency of each allele from the selected polymorphic loci to determine a percent fetal cell-free DNA in the sample, quantifying a relative frequency of loci from the first target genomic region and a relative frequency of loci from the second target genomic region, and computing a likelihood of the presence or absence of a fetal aneuploidy using the relative frequency of the loci from the first and second target genomic regions and the percent fetal cell-free DNA to determine the likelihood of the presence or absence of a fetal aneuploidy.

[00046] In certain aspects, the invention further comprises comparing the relative frequency of the loci from the first and second target genomic regions and adjusting the relative frequency of the loci from the first and second target genomic regions based on the percent fetal cell-free DNA to determine the likelihood of the presence or absence of a fetal aneuploidy. In specific aspects, the relative frequencies of each selected locus for each target genomic region are summed and the sums for each chromosome are compared to calculate a target genomic region ratio.

[00047] The percent fetal cell free DNA of a sample can be calculated by detecting levels of one or more non-maternal contributed loci, *e.g.,* non-maternal loci on the Y-chromosome and/or non-maternal loci are autosomal loci. In preferred aspects, the non-maternal loci comprise one or more genetic variations compared to maternal loci, *e.g.,* SNPs or methylation differences.

[00048] In certain embodiments, the ligation products are cleaved (*e.g.*, using enzymatic cleaving mechanisms such as a restriction endonuclease) to reduce the size of the ligation product while leaving the capture region and label binding region available for detection. In certain aspects, the cleavage occurs after the universal amplification.

[00049] In preferred embodiments, the loci and fixed sequence oligonucleotides hybridized to the loci are isolated from unbound fixed sequence oligonucleotides following hybridization to remove excess unbound oligonucleotides in the reaction; *e.g.,* through a washing step or enzymatic degradation of the unbound oligonucleotides.

[00050] The first and second sets of fixed sequence oligonucleotides used in the methods preferably comprise—in addition to at least one capture region—universal primer regions that may be used to amplify the ligation products. Alternatively, universal primer sequences may be added to the ends of the ligation products following ligation, *e.g.,* through the introduction of adapters comprising universal primer sequences.

[00051] In certain aspects, the fixed sequence oligonucleotides of the invention comprise one or more indices. These indices may serve, in addition to the capture regions, as surrogate

sequences to identify the loci, or a particular allele of a locus. In particular, these indices may serve as surrogate identification sequences to detect hybridization of the ligation product or amplicons thereof to an array. In specific methods, the first or second fixed sequence oligonucleotide in each set of fixed sequence oligonucleotides comprises an allele index that associates a specific allele with the fixed sequence oligonucleotide.

[00052] In certain specific aspects, the method is carried out for at least 50 loci from each target genomic region, more preferably between 50-100 loci, more preferably between 100-200 loci, more preferably between 200-500 loci, more preferably between 500-1000 loci, preferably between 1000-2000 loci, preferably between 2000-5000 loci, and preferably between 5000-10,000 loci from a target genomic region, or any intervening range therein. In certain aspects, in addition to the target genomic regions, at least 50 selected polymorphic loci are interrogated. More preferably, between 50-100 selected polymorphic loci are interrogated, more preferably between 100-200 selected polymorphic loci, more between 200-500 selected polymorphic loci, more between 500-1000 selected polymorphic loci, between 1000-2000 selected polymorphic loci, between 2000-5000 selected polymorphic loci, and between 5000-10,000 selected polymorphic loci are interrogated, including all intervening ranges.

[00053] In other aspects, the assay methods are estimated to detect at least 5 capture regions corresponding to each locus within a target genomic region, more preferably at least 10 capture regions corresponding to each locus within a target genomic region, more preferably at least 20 capture regions corresponding to each locus within a target genomic region, preferably at least 50 capture regions corresponding to each locus within a target genomic region, more preferably at least 100 capture regions corresponding to each locus within a target genomic region, more preferably at least 200 capture regions corresponding to each locus within a target genomic region. In some embodiments, no more than 5000 capture regions corresponding to each locus within a target genomic region are detected for each sample. In other embodiments

no more than 2000 capture regions corresponding to each locus within a target genomic region are detected for each sample.

[00054] These aspects and other features and advantages of the invention are described in more detail below.

BRIEF DESCRIPTION OF THE FIGURES

[00055] FIG. 1 is a flow chart describing one general aspect of the invention.

[00056] FIG. 2 illustrates an embodiment of a method of the invention that utilizes hybridization detection of loci.

[00057] FIG. 3 illustrates an alternative embodiment of a method of the invention that utilizes hybridization detection of loci.

[00058] FIG. 4 illustrates another alternative embodiment of a method of the invention that utilizes hybridization detection of loci.

[00059] FIG, 5 illustrates yet another alternative embodiment of a method of the invention that utilizes hybridization detection of loci.

[00060] FIG. 6 illustrates another alternative embodiment of a method of the invention that utilizes bridging oligonucleotides in combination with fixed sequence oligonucleotides and hybridization detection of loci.

[00061] FIG. 7 illustrates another alternative embodiment of a method of the invention that utilizes hybridization detection of loci to detect polymorphisms.

[00062] FIG. 8 illustrates another alternative embodiment of a method of the invention that utilizes hybridization detection of loci to detect polymorphisms.

[00063] FIG. 9 illustrates another alternative embodiment of a method of the invention that utilizes hybridization detection of nucleic acid regions to detect polymorphisms.

[00064] FIG. 10 illustrates a method of the invention that utilizes hybridization detection of nucleic acid regions with a bridging oligonucleotide and dual cleavage.

[00065] FIG. 11 of a method of the invention that utilizes hybridization detection of nucleic acid regions with a bridging oligonucleotide and dual cleavage to detect polymorphisms.

[00066] FIG. 12 illustrates a method of the invention that utilizes hybridization detection of nucleic acid regions resulting from a single cleavage event and employing differentially labeled universal primers.

[00067] FIG. 13 illustrates an alternative method to that illustrated in Figure 12 also utilizing hybridization detection of nucleic acid regions resulting from a single cleavage event and employing differentially labeled universal primers.

[00068] FIG. 14 shows the distribution of assay variability across samples for arrays and next generation sequencing.

DEFINITIONS

[00069] The terms used herein are intended to have the plain and ordinary meaning as understood by those of ordinary skill in the art. The following definitions are intended to aid the reader in understanding the present invention, but are not intended to vary or otherwise limit the meaning of such terms unless specifically indicated.

[00070] The term “allele index” refers generally to a series of nucleotides that corresponds to a specific SNP. The allele index may contain additional nucleotides that allow for the detection of deletion, substitution, or insertion of one or more bases. The allele index may be combined with any other index to create one index that provides information for two properties (*e.g.,* sample-identification index, allele-locus index).

[00071] "Array" refers to a solid phase support having a surface, preferably but not exclusively a planar or substantially planar surface, which carries an array of sites containing nucleic acids such that each site of the array comprises substantially identical or identical copies of oligonucleotides or polynucleotides and is spatially defined and not overlapping with other member sites of the array; that is, the sites are spatially discrete. The array or microarray can

also comprise a non-planar interrogatable structure with a surface such as a bead or a well. The oligonucleotides or polynucleotides of the array may be covalently bound to the solid support, or may be non-covalently bound. Conventional microarray technology is reviewed in, *e.g.*, Schena, Ed., *Microarrays: A Practical Approach*, IRL Press, Oxford (2000). "Array analysis", "analysis by array" or "analysis by microarray" refers to analysis, such as, *e.g.*, sequence analysis, of one or more biological molecules using an array. The term array refers to any format of arrayed solid substrates, including a microarray, arrayed beads, an array of molecules within wells, or "liquid" arrays.

[00072] The term "binding pair" means any two molecules that specifically bind to one another using covalent and/or non-covalent binding, and which can be used, e.g., for attachment of genetic material to a substrate. Examples include, but are not limited to, ligands and their protein binding partners, *e.g.,* biotin and avidin, biotin and streptavidin, an antibody and its particular epitope, and the like.

[00073] The term "chromosomal abnormality" refers to any genetic variant for all or part of a chromosome. The genetic variants may include but are not limited to any copy number variant such as duplications or deletions, translocations, inversions, and mutations.

[00074] The terms "complementary" or "complementarity" are used in reference to nucleic acid molecules (*i.e.*, a sequence of nucleotides) that are related by base-pairing rules. Complementary nucleotides are, generally, A and T (or A and U), or C and G. Two single stranded RNA or DNA molecules are said to be substantially complementary when the nucleotides of one strand, optimally aligned and with appropriate nucleotide insertions or deletions, pair with at least about 90% to about 95% complementarity, and more preferably from about 98% to about 100% complementarity, and even more preferably with 100% complementarity. Alternatively, substantial complementarity exists when an RNA or DNA strand will hybridize under selective hybridization conditions to its complement. Selective hybridization conditions include, but are not limited to, stringent hybridization conditions. Stringent

hybridization conditions will typically include salt concentrations of less than about 1 M, more usually less than about 500 mM and preferably less than about 200 mM. Hybridization temperatures are generally at least about 2°C to about 6°C lower than melting temperatures (T_m).

[00075] The term "diagnostic tool" as used herein refers to any composition or system of the invention used in combination as, for example, in a system in order to carry out a diagnostic test or detection system on a patient sample.

[00076] The term "hybridization" generally means the reaction by which the pairing of complementary strands of nucleic acid occurs. DNA is usually double-stranded, and when the strands are separated they will re-hybridize under the appropriate conditions. Hybrids can form between DNA-DNA, DNA-RNA or RNA-RNA. They can form between a short strand and a long strand containing a region complementary to the short one. Imperfect hybrids can also form, but the more imperfect they are, the less stable they will be (and the less likely to form).

[00077] As used herein the term "ligase" refers generally to a class of enzymes, DNA ligases (typically T4 DNA ligase), which can link pieces of DNA together. The pieces must have compatible ends–either with both of them blunt or with mutually-compatible sticky ends–and the reaction requires ATP. "Ligation" is the process of joining two pieces of DNA together.

[00078] The terms "locus" and "loci" as used herein refer to a nucleic acid region of known location in a genome.

[00079] The term "maternal sample" as used herein refers to any sample taken from a pregnant mammal which comprises both fetal and maternal cell-free DNA. Preferably, maternal samples for use in the invention are obtained through relatively non-invasive means, *e.g.,* phlebotomy or other standard techniques for extracting peripheral samples from a subject.

[00080] The term "oligonucleotides" or "oligos" as used herein refers to linear oligomers of natural or modified nucleic acid monomers, including deoxyribonucleotides, ribonucleotides, anomeric forms thereof, peptide nucleic acid monomers (PNAs), locked nucleotide acid

monomers (LNA), and the like, or a combination thereof, capable of specifically binding to a single-stranded polynucleotide by way of a regular pattern of monomer-to-monomer interactions, such as Watson-Crick type of base pairing, base stacking, Hoogsteen or reverse Hoogsteen types of base pairing, or the like. Usually monomers are linked by phosphodiester bonds or analogs thereof to form oligonucleotides ranging in size from a few monomeric units, *e.g.*, 8-12, to several tens of monomeric units, *e.g.*, 100-200 or more. Suitable nucleic acid molecules may be prepared by the phosphoramidite method described by Beaucage and Carruthers (Tetrahedron Lett., 22:1859-1862 (1981)), or by the triester method according to Matteucci, et al. (J. Am. Chem. Soc., 103:3185 (1981)), both of which are incorporated herein by reference, or by other chemical methods such as using a commercial automated oligonucleotide synthesizer.

[00081] As used herein "nucleotide" refers to a base-sugar-phosphate combination. Nucleotides are monomeric units of a nucleic acid sequence (DNA and RNA). The term nucleotide includes ribonucleoside triphosphates ATP, UTP, CTG, GTP and deoxyribonucleoside triphosphates such as dATP, dCTP, dITP, dUTP, dGTP, dTTP, or derivatives thereof. Such derivatives include, for example, [αS]dATP, 7-deaza-dGTP and 7-deaza-dATP, and nucleotide derivatives that confer nuclease resistance on the nucleic acid molecule containing them. The term nucleotide as used herein also refers to dideoxyribonucleoside triphosphates (ddNTPs) and their derivatives. Illustrated examples of dideoxyribonucleoside triphosphates include, but are not limited to, ddATP, ddCTP, ddGTP, ddITP, and ddTTP.

[00082] As used herein the term "polymerase" refers to an enzyme that links individual nucleotides together into a long strand, using another strand as a template. There are two general types of polymerase—DNA polymerases, which synthesize DNA, and RNA polymerases, which synthesize RNA. Within these two classes, there are numerous sub-types of polymerases, depending on what type of nucleic acid can function as template and what type of nucleic acid is formed.

[00083] As used herein "polymerase chain reaction" or "PCR" refers to a technique for replicating a specific piece of target DNA *in vitro*, even in the presence of excess non-specific DNA. Primers are added to the target DNA, where the primers initiate the copying of the target DNA using nucleotides and, typically, Taq polymerase or the like. By cycling the temperature, the target DNA is repetitively denatured and copied. A single copy of the target DNA, even if mixed in with other, random DNA, can be amplified to obtain billions of replicates. The polymerase chain reaction can be used to detect and measure very small amounts of DNA and to create customized pieces of DNA. In some instances, linear amplification methods may be used as an alternative to PCR.

[00084] The term "polymorphism" as used herein refers to any genetic changes or variants in a loci that may be indicative of that particular loci, including but not limited to single nucleotide polymorphisms (SNPs), methylation differences, short tandem repeats (STRs), and the like.

[00085] Generally, a "primer" is an oligonucleotide used to, *e.g.,* prime DNA extension, ligation and/or synthesis, such as in the synthesis step of the polymerase chain reaction or in the primer extension techniques. A primer may also be used in hybridization techniques as a means to provide complementarity of a nucleic acid region to a capture oligonucleotide for detection of a specific nucleic acid region.

[00086] The term "research tool" as used herein refers to any composition or system of the invention used for scientific enquiry, academic or commercial in nature, including the development of pharmaceutical and/or biological therapeutics. The research tools of the invention are not intended to be therapeutic or to be subject to regulatory approval; rather, the research tools of the invention are intended to facilitate research and aid in such development activities, including any activities performed with the intention to produce information to support a regulatory submission.

[00087] The term "sample" refers to any sample comprising all or a portion of the genetic information of an organism, including but not limited to virus, bacteria, fungus, plants and

animals, and in particular mammals. The genetic information that can be interrogated within a genetic sample includes genomic DNA (both coding and non-coding regions), mitochondrial DNA, RNA, and nucleic acid products derived from each of these. Such nucleic acid products include cDNA created from mRNA or products of pre-amplification to increase the material for analysis.

[00088] The term "target genomic region" refers to all or a portion of a chromosome or chromosomes, including complete chromosomes, sub-chromosomal regions, groups of loci and single loci.

DETAILED DESCRIPTION OF THE INVENTION

[00089] The practice of the techniques described herein may employ, unless otherwise indicated, conventional techniques and descriptions of organic chemistry, polymer technology, molecular biology (including recombinant techniques), cell biology, biochemistry, and array technology, which are within the skill of those who practice in the art. Such conventional techniques include polymer array synthesis, hybridization and ligation of polynucleotides, and detection of hybridization using a label. Specific illustrations of suitable techniques can be had by reference to the examples herein. However, other equivalent conventional procedures can, of course, also be used. Such conventional techniques and descriptions can be found in standard laboratory manuals such as Green, et al., Eds. (1999), *Genome Analysis: A Laboratory Manual Series* (Vols. I-IV); Weiner, Gabriel, Stephens, Eds. (2007), *Genetic Variation: A Laboratory Manual*; Dieffenbach, Dveksler, Eds. (2003), *PCR Primer: A Laboratory Manual*; Bowtell and Sambrook (2003), *DNA Microarrays: A Molecular Cloning Manual*; Mount (2004), *Bioinformatics: Sequence and Genome Analysis*; Sambrook and Russell (2006), *Condensed Protocols from Molecular Cloning: A Laboratory Manual*; and Sambrook and Russell (2002), *Molecular Cloning: A Laboratory Manual* (all from Cold Spring Harbor Laboratory Press); Stryer, L. (1995) *Biochemistry* (4th Ed.) W.H. Freeman, New York N.Y.; Gait, "*Oligonucleotide Synthesis: A Practical Approach*" 1984, IRL Press, London; Nelson and Cox (2000), *Lehninger,*

Principles of Biochemistry 3rd Ed., W. H. Freeman Pub., New York, N.Y.; and Berg et al. (2002) *Biochemistry,* 5th Ed., W.H. Freeman Pub., New York, N.Y., all of which are herein incorporated in their entirety by reference for all purposes.

[00090] Note that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to “an allele” refers to one or more copies of allele with various sequence variations, and reference to “the detection system” includes reference to equivalent steps and methods known to those skilled in the art, and so forth.

[00091] Unless defined otherwise, 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. All publications mentioned herein are incorporated by reference in their entirety for all purposes, including the purpose of describing and disclosing devices, reagents, techniques and methodologies that may be used in or in connection with the presently described invention.

[00092] Where a range of values is provided, it is understood that each intervening value, between the upper and lower limit of that range and any other stated or intervening value in that stated range is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either both of those included limits are also included in the invention.

[00093] In the following description, numerous specific details are set forth to provide a more thorough understanding of the present invention. However, it will be apparent to one of skill in the art that the present invention may be practiced without one or more of these specific details. In other instances, well-known features and procedures well known to those skilled in the art have not been described in order to avoid obscuring the invention.

The Invention in General

[00094] The invention provides assay methods to identify copy number variants of nucleic acid regions (including loci, sets of loci and larger target genomic regions, e.g., chromosomes), including insertions, deletions, translocations, mutations and polymorphisms in a genetic sample. In one aspect, the assay methods interrogate loci from two or more target genomic regions in a sample using a directed ligation assay followed by detection of labelled oligonucleotides attached to an array. Quantification of the labelled oligonucleotides allows determination of an atypical copy number of a particular target genomic region based on a comparison between the quantities of detected loci from the target genomic regions (e.g., comparison between two or more portions of a single chromosome or comparison between two or more different chromosomes) in the sample or by comparison to a reference chromosome from the same or a different sample.

[00095] In some embodiments, the method employs directed analysis of target genomic regions in a sample using sets of fixed sequence oligonucleotides that selectively hybridize to loci within two or more target genomic regions. The fixed sequence oligonucleotides are directly or indirectly ligated to create ligation products. The ligation products corresponding to loci associated with a first target genomic region are associated with a first detectable label and ligation products corresponding to loci associated with a second target genomic region are associated with a second detectable label. If the first and second detectable labels are quantified, the relative frequency of each of the first and second target genomic regions can be determined. In certain aspects of the invention, the method employs two different labels that are used to identify two different target genomic regions. In other aspects of the invention, the method employs three different labels corresponding to three different target genomic regions, and so on.

[00096] The ligation products are detected by hybridization, and in particular by hybridization to an array of capture probes complementary to capture regions present in the ligation products.

In certain embodiments, the ligation products are detected using "universal arrays" that comprise features having the same or substantially similar capture probes. In certain other embodiments, the arrays comprises two or more sets of multiple features with a common sequence, with each set having a different sequence, *e.g.,* an array where up to hundreds of the features on the array have substantially the same sequence. In either case, the capture probes on the array are complementary to the capture regions of the ligation products rather than to the sequence of the loci or their complements. These arrays can be used to interrogate any loci for any target genomic region(s) regardless of the sequence of the loci.

[00097] The capture regions are preferably introduced to the ligation products in the fixed sequence oligonucleotides that are used to interrogate the loci in the sample. In some preferred embodiments, the capture regions are the same amongst all fixed sequence oligonucleotides used, so that ligation products or amplicons or cleavage products thereof from all loci hybridize competitively to capture probes of the same sequence on the array. In other embodiments, the array is comprised of many different capture probes, and the sets of fixed sequence oligonucleotides from different loci comprise different capture regions.

[00098] Fig. 1 is a flow chart 100 illustrating an exemplary method of the invention. In step 102, a sample is provided. In step 104, sets of fixed sequence oligonucleotides comprising a label binding region are introduced to the sample under conditions that allow the fixed sequence oligonucleotides to hybridize to loci in target genomic regions, and in step 106 the oligonucleotides are hybridized to the target genomic regions. In step 108, the hybridized fixed sequence oligonucleotides from each set are ligated to create ligation products which are then amplified in step 110 to produce amplicons complementary to the ligation products. In step 112, the amplicons are introduced to a hybridization array and allowed to hybridize competitively to capture probes on the array. In step 114, a set of labelled oligonucleotides are introduced to the amplicons and allowed to hybridize to complementary sequences on the amplicons. Optionally,

the labelled oligonucleotides are ligated to the capture probe on the array (not shown). In step 116, the labels are detected.

[00099] Each fixed sequence oligonucleotide of each set comprises a region complementary to a selected locus (as described in more detail in Figure 2). At least one fixed sequence oligonucleotide of each set further comprises a capture region, which may be the same for all sets of fixed sequence oligonucleotides used to interrogate two or more target genomic regions, may be the same for pairs of sets of fixed sequence oligonucleotides used to interrogate two or more target genomic regions, or may be different between sets of fixed sequence oligonucleotides for individual target genomic regions. Additionally, depending on the embodiment, one fixed sequence oligonucleotide of a set comprises either a detectable label or a label binding region for association of the fixed sequence oligonucleotide with the detectable label. In specific embodiments, the label binding region can be a region complementary to a labeled oligonucleotide associated with a detectable label.

[000100] In some embodiments, the fixed sequence oligonucleotide of each set that comprises the capture region will not comprise the label or label binding region; that is, the other fixed sequence oligonucleotide of the set comprises the label or label binding region (see, *e.g.,* exemplary embodiments illustrated in Figs. 2, 3 and 5-8); in other embodiments, the fixed sequence oligonucleotide of each set that comprises the capture region also will comprise the label or label binding region (see, *e.g.,* exemplary embodiment illustrated in Fig. 4).

[000101] In certain specific aspects, a first set of fixed sequence oligonucleotides hybridizes to loci in a first target genomic region while a second set of fixed sequence oligonucleotides hybridizes to loci in a second target genomic region. After ligation to produce ligation products, the ligation products are optionally amplified using universal primers, and then are hybridized to an array. In other embodiments, the amplification product is cleaved (*e.g.,* using a restriction endonuclease) and a portion of the amplification product comprising the capture region is introduced to the array for hybridization and detection. In preferred embodiments, each set of

fixed sequence oligonucleotides used to interrogate a target genomic region contains the same label or label binding region. That is, all of the fixed sequence oligonucleotides of the first set are associated with a first label, all of the fixed sequence oligonucleotides of the second set are associated with a second label, and all of the fixed sequence oligonucleotides in a third set are associated with a third label.

[000102] If the fixed sequence oligonucleotides are labeled directly, the ligation products, amplicons or cleavage products thereof can be hybridized to the capture probes on the array and detected by readout from the labels. If the fixed sequence oligonucleotides instead contain a label binding region that is complementary to a labeled oligonucleotide (a "label binding sequence"), a labeled oligonucleotide must be added to the ligation product or amplicons before detection. In either scenario, the labels are then detected and quantified and the relative frequency of each label determined. Quantifying each label allows for quantification of each target genomic region.

[000103] In some embodiments, all sets of first and second fixed sequence oligonucleotides contain substantially the same capture region. In these embodiments, because ligation products from all loci from all target genomic regions share the same capture region complementary to the capture probes on the array, the ligation products from each target genomic region compete to hybridize to the capture probes on a universal array.

[000104] In other embodiments, the capture probes on an array comprise multiple sequences complementary to different capture regions, and the array comprises features that contain these different capture probes. In some such embodiments, each capture probe may hybridize to an unique capture region. In other such embodiments, more than one capture probe, representing loci from different genomic regions, may hybridize to a single capture region.

[000105] In other embodiments, different loci from the same or different target genomic regions may be configured to competitively hybridize against one another and thus would

comprise the same capture region, while other loci from the same or different target genomic regions may be configured to competitively hybridize against one another, depending on the assay.

[000106] The target genomic regions may be large genomic regions, such as whole chromosomes, or may be smaller genomic regions such as sub-regions of a single chromosome or sub-regions on different chromosomes, even down to a single locus. Thus, the invention may be used to detect genomic variations such as aneuploidies and partial aneuploidies, as well as mutations, SNPs, rearrangements, insertions and deletions. In the case where whole chromosomes are compared, the first target genomic region may be, e.g., chromosome 21, and all loci to be interrogated with the first set of fixed sequence oligonucleotides will be from chromosome 21.

[000107] In the ligation assay, if the fixed sequence oligonucleotides bind to immediately adjacent regions in a selected locus, the fixed oligonucleotides may be ligated to create ligation products which are associated with target genomic region-specific labels. In the case where fixed sequence oligonucleotides do not bind to immediately adjacent regions within the genomic region—i.e., there is a gap between the hybridized fixed sequence oligonucleotides—the gap can be closed using primer extension, and/or one or more bridging oligonucleotides. Once the oligonucleotides are hybridized contiguously, either directly or following an extension operation or introduction of a bridging oligonucleotide, they may then be ligated to create ligation products which are associated with target genomic region-specific labels.

[000108] In certain aspects, a first set of fixed sequence oligonucleotides are selective for a first chromosome or first target genomic region and a second set of fixed sequence oligonucleotides are selective for a different chromosome or second target genomic region. Fig. 2 illustrates one embodiment in which each set of labeled fixed sequence oligonucleotides hybridize to loci on different chromosomes and ligation products are evaluated competitively on an array comprising capture probes. Two sets of labeled fixed sequence oligonucleotides 202,

204 are provided, each set having a first fixed sequence oligonucleotide 206, 208 comprising a sequence that is complementary to a selected locus 210, 212 and a label 214, 216 and a second fixed sequence oligonucleotide 218, 220 comprising a sequence complementary to the selected locus 222, 224 and a capture region 226, 228. The labels 214, 216 are different for each set of fixed sequence oligonucleotides to allow differentiation between the first target genomic region (in this case, a first chromosome) and the second target genomic region (in this case, a second chromosome) during detection. In step 230, the sets of fixed sequence oligonucleotides 202, 204 are introduced to a sample and allowed to hybridize to loci 232, 234 on two different chromosomes. Following hybridization, the unhybridized fixed sequence oligonucleotides preferably are separated from the remainder of the sample (not shown). In step 236, the fixed sequence oligonucleotides are ligated to create ligation products 238, 240 comprising capture regions 226, 228 and labels 214, 216. Although the fixed sequence oligonucleotides are illustrated in FIG. 2 as being hybridized adjacently in the loci, there may also be a gap that can be filled, *e.g.* using an extension reaction or using a bridging oligonucleotide that hybridizes adjacently between the fixed sequence oligonucleotides. In step 242, the ligation products 238, 240 are introduced to a hybridization array 244 comprising a plurality of capture probes 246 wherein the capture regions 226, 228 of the ligation products 238, 240 competitively hybridize to the capture probes 246. In a preferred embodiment, 226 and 228 have substantially the same sequence. Following hybridization of the ligation products to the array, unhybridized ligation products preferably are removed from the array (not shown). The labels 214, 216 can then be detected using an appropriate detection mechanism depending on the type of label used and the loci corresponding to each of the first and second chromosomes can be quantified to determine the presence and amount of each chromosome in the genetic sample.

[000109] According to the present invention, a “nucleotide” may be unlabeled or detectably labeled by well-known techniques. Fluorescent labels and their attachment to oligonucleotides

are described in many reviews, including Haugland, *Handbook of Fluorescent Probes and Research Chemicals*, 9th Ed., Molecular Probes, Inc., Eugene OR (2002); Keller and Manak, *DNA Probes*, 2nd Ed., Stockton Press, New York (1993); Eckstein, Ed., *Oligonucleotides and Analogues: A Practical Approach*, IRL Press, Oxford (1991); Wetmur, *Critical Reviews in Biochemistry and Molecular Biology*, 26:227-259 (1991); and the like. Other methodologies applicable to the invention are disclosed in the following sample of references: Fung et al., U.S. Pat. No. 4,757,141; Hobbs, Jr., et al., U.S. Pat. No. 5,151,507; Cruickshank, U.S. Pat. No. 5,091,519; Menchen et al., U.S. Pat. No. 5,188,934; Begot et al., U.S. Pat. No. 5,366,860; Lee et al., U.S. Pat. No. 5,847,162; Khanna et al., U.S. Pat. No. 4,318,846; Lee et al., U.S. Pat. No. 5,800,996; Lee et al., U.S. Pat. No. 5,066,580: Mathies et al., U.S. Pat. No. 5,688,648; and the like. Labeling can also be carried out with quantum dots, as disclosed in the following patents and patent publications: U.S. Pat. Nos. 6,322,901; 6,576,291; 6,423,551; 6,251,303; 6,319,426; 6,426,513; 6,444,143; 5,990,479; 6,207,392; 2002/0045045; and 2003/0017264. Detectable labels include, for example, radioactive isotopes, fluorescent labels, chemiluminescent labels, bioluminescent labels and enzyme labels. Fluorescent labels of nucleotides may include but are not limited fluorescein, 5-carboxyfluorescein (FAM), 2'7'-dimethoxy-4'5-dichloro-6-carboxyfluorescein (JOE), rhodamine, 6-carboxyrhodamine (R6G), N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA), 6-carboxy-X-rhodamine (ROX), 4-(4'dimethylaminophenylazo) benzoic acid (DABCYL), CASCADE BLUE® (pyrenyloxytrisulfonic acid), OREGON GREEN™ (2',7'-difluorofluorescein), TEXAS RED™ (sulforhodamine 101 acid chloride), Cyanine and 5-(2'-aminoethyl)aminonaphthalene-1-sulfonic acid (EDANS). Specific examples of fluroescently labeled nucleotides include [R6G]dUTP, [TAMRA]dUTP, [R110]dCTP, [R6G]dCTP, [TAMRA]dCTP, [JOE]ddATP, [R6G]ddATP, [FAM]ddCTP, [R110]ddCTP, [TAMRA]ddGTP, [ROX]ddTTP, [dR6G]ddATP, [dR110]ddCTP, [dTAMRA]ddGTP, and [dROX]ddTTP available from Perkin Elmer, Foster City, Calif. FluoroLink DeoxyNucleotides, FluoroLink Cy3-dCTP, FluoroLink Cy5-dCTP, FluoroLink FluorX-dCTP, FluoroLink Cy3-dUTP, and FluoroLink Cy5-

dUTP available from Amersham, Arlington Heights, IL; Fluorescein-15-dATP, Fluorescein-12-dUTP, Tetramethyl-rodamine-6-dUTP, IR770-9-dATP, Fluorescein-12-ddUTP, Fluorescein-12-UTP, and Fluorescein-15-2'-dATP available from Boehringer Mannheim, Indianapolis, IN; and Chromosomee Labeled Nucleotides, BODIPY-FL-14-UTP, BODIPY-FL-4-UTP, BODIPY-TMR-14-UTP, BODIPY-TMR-14-dUTP, BODIPY-TR-14-UTP, BODIPY-TR-14-dUTP, CASCADE BLUE®-7-UTP (pyrenyloxytrisulfonic acid-7-UTP), CASCADE BLUE®-7-dUTP (pyrenyloxytrisulfonic acid-7-dUTP), fluorescein-12-UTP, fluorescein-12-dUTP, OREGON GREEN™ 488-5-dUTP (2',7'-difluorofluorescein-5-dUTP), RHODAMINE GREEN™-5-UTP ((5-{2-[4-(aminomethyl)phenyl]-5-(pyridin-4-yl)-1H-i-5-UTP)), RHODAMINE GREEN™-5-dUTP ((5-{2-[4-(aminomethyl)phenyl]-5-(pyridin-4-yl)-1H-i-5-dUTP)), tetramethylrhodamine-6-UTP, tetramethylrhodamine-6-dUTP, TEXAS RED™-5-UTP (sulforhodamine 101 acid chloride-5-UTP), TEXAS RED™-5-dUTP (sulforhodamine 101 acid chloride-5-dUTP), and TEXAS RED™-12-dUTP (sulforhodamine 101 acid chloride-12-dUTP) available from Molecular Probes, Eugene, OR.

[000110] In certain aspects of the invention, nucleic acids from the sample are associated with a substrate—*e.g.*, using binding pairs such as, *e.g.,* biotin and streptavidin, to attach the genetic material to a substrate surface or direct covalent attachment—before adding the sets of fixed sequence oligonucleotides to the sample. Briefly, a first member of a binding pair (*e.g.*, biotin) can be associated with nucleic acids from the sample, and the nucleic acids attached to a substrate via a second member of the binding pair (e.g., avidin or streptavidin) on the surface of the substrate. Attachment of the nucleic acids from the sample can be particularly useful in removing unhybridized oligonucleotides following hybridization of the fixed sequence oligonucleotides and/or the bridging oligonucleotides to the loci. Briefly, the nucleic acids from the sample can be hybridized to the fixed sequence oligonucleotides, and then the hybridization complexes are subsequently bound to a substrate. Alternatively, the nucleic acids from the sample can be attached to a solid support prior to hybridization of the fixed sequence

oligonucleotides or at the same time. Either way, following hybridization and attachment of the nucleic acids to a solid support, or alternatively following ligation of the hybridized oligonucleotides, the surface of the support can be treated to remove any unhybridized or unligated oligonucleotides, *e.g.,* by washing or other removal methods such as degradation of oligonucleotides as discussed in Willis et al., U.S. Pat. Nos. 7,700,323 and 6,858,412. Degradation of the oligonucleotides is a preferred aspect when the two fixed sequence oligonucleotides are on the same probe such that ligation results in a circularized probe. Exonucleases may then be used to degrade non-circularized nucleic acids, including excess probes and sample DNA.

[000111] There are a number of methods that may be used to associate nucleic acids with binding pairs. For example, numerous methods may be used for labeling nucleic acids with biotin, including random photobiotinylation, end-labeling with biotin, replicating with biotinylated nucleotides, and replicating with a biotin-labeled primer.

[000112] The number of loci analyzed for each chromosome in the methods of the invention may vary from two to 20,000 or more per target genomic region analyzed. In a preferred aspect, the number of loci per target genomic region is between 48 and 1000. In another aspect, the number of loci per target genomic region is at least 100. In another aspect, the number of loci per target genomic region is at least 400. In another aspect, the number of loci per target genomic region is no more than 1000. In another aspect, the number of loci per target genomic region is at least 500 but no more than 2000.

[000113] While the embodiment illustrated in FIG. 2 uses fixed sequence oligonucleotides coupled directly to a detectable label, a label may instead be provided by a separate, labeled oligonucleotide that is hybridized to the ligation products of the fixed sequence oligonucleotides or amplicons or cleaved amplicon (as described below) thereof, to allow detection. Optionally, the labeled oligonucleotide is ligated to the capture probe following hybridization to the fixed sequence oligonucleotides or amplicons or cleaved amplicon. Fig. 3 is an illustration of one

embodiment of the invention in which the fixed sequence oligonucleotides are hybridized to loci of interest, ligated, amplified and introduced to an array prior to hybridization of a labeled oligonucleotide and detection of the label. In the method depicted in FIG. 3, two sets of fixed sequence oligonucleotides 302, 304 are provided, wherein the sets comprise a first fixed sequence oligonucleotide 306, 308 each comprising sequences complementary to a selected locus 310, 312, label binding region 314, 316 and universal primer regions 318, 320; and a second fixed sequence oligonucleotide 322, 324 comprising sequences complementary to the selected locus 326, 328, a capture region 330, 332 and a universal primer region 334, 336. In many embodiments, the capture region 330, 332 comprise substantially the same sequence and will both hybridize to the same capture probe on an array. The label binding regions 314, 316 comprise sequences that are different for each set of fixed sequence oligonucleotides 302, 304 allowing differential labeling of the fixed sequence oligonucleotides associated with the loci for each target genomic region, while the capture regions 330, 332 are the same for both sets of fixed sequence oligonucleotides 302, 304 to allow for competitive hybridization of the ligation products to capture features on an array. In step 338, the sets of fixed sequence oligonucleotides 302, 304 are introduced to a sample and allowed to hybridize to loci 340, 342 of different target genomic regions. Following hybridization, unhybridized fixed sequence oligonucleotides preferably are separated from the remainder of the genetic sample (not shown).

[000114] In step 344, the sets of fixed sequence oligonucleotides 302, 304 are ligated to create ligation products 346, 348. In step 358, universal primers 350, 352, 354 and 356 are introduced to the ligation products 346, 348 which bind to the universal primer regions 318, 334, 320, and 336, respectively, and create amplicons 360, 362 each comprising capture regions 364, 366 and label binding regions 368, 370. In certain preferred embodiments, 350 and 354 have substantially the same sequence, which is complementary to both 318 and 320, and 352 and 356 have substantially the same sequence, which is complementary to both 334 and 336.

The amplicons are introduced to a hybridization array 372 comprising a plurality of capture probes 374 wherein the capture regions 364, 366 of the amplicons 360, 362 hybridize to the capture probes 374. In step 378 labeled oligonucleotides 380, 382 are introduced to the array 372 where the label binding regions 368, 370 of the amplicons 360, 362 hybridize to target recognition regions 384, 386 of the labeled oligonucleotides 380, 382. Optionally, the labelled oligonucleotides are ligated to the capture probe on the array (not shown). Following hybridization of the labeled oligonucleotides, unhybridized labeled oligonucleotides preferably are separated from the array (not shown). The labels can then be detected and the loci corresponding to each target genomic region quantified to provide information on the presence and quantity of each target genomic region in the sample.

[000115] In certain embodiments, such as the embodiment shown in FIG. 3, labeled oligonucleotides are hybridized to the fixed sequence oligonucleotides, or amplicons or cleaved amplicons thereof, after the ligation products or amplicons or cleaved amplicons are hybridized to an array. In other certain embodiments, the labeled oligonucleotides are hybridized to the fixed sequence oligonucleotides, or amplicons or cleaved amplicons thereof, prior to hybridization to an array.

[000116] To facilitate hybridization of ligation products or amplicons thereof, the size of the ligation products or amplicons may be reduced prior to hybridization to an array. In certain embodiments, the ligation products are cleaved (e.g., using restriction endonucleases or other enzymatic cleaving mechanisms) to reduce the size of the ligation product to be detected, leaving, e.g., the capture region and the label binding region available for detection. Detection of a cleaved labeled ligation product or a cleaved amplicon serves as a surrogate in lieu of detecting the entire ligation product.

[000117] Reducing the size of the ligation products, amplicons and/or labeled ligation products can facilitate binding on the array, *e.g.*, by improving hybridization kinetics and by decreasing steric hindrance. In certain embodiments, reduction in the size of the ligation

products is accomplished by cleaving the ligation products or amplicons using a restriction enzyme. For example, in certain embodiments, one of the fixed sequence oligonucleotides in each set of fixed sequence oligonucleotides comprises a restriction enzyme recognition site proximal to the capture region or label binding region (or the corresponding complementary sequences thereof depending on the embodiment). A restriction enzyme can be used to cleave the ligation product at the restriction enzyme recognition site leaving the label or label binding region and the capture region available for hybridization and detection. A restriction enzyme recognition site may be located in any position that leaves the capture region and the label binding region available for detection after cleavage. For example, the restriction enzyme recognition site may be located directly next to the capture region or the label binding region or within a few bases from either of the capture region or label binding region. Preferably cleavage is carried out prior to hybridization of the labeled oligonucleotide to the ligation products or amplicons. In certain other aspects, cleavage occurs after hybridization to the array for detection.

[000118] FIG. 4 is an illustration of a specific embodiment of the invention in which the ligation product is cleaved prior to hybridization to an array. In the method depicted in FIG. 4, two sets of fixed sequence oligonucleotides 402, 404 are provided. Each set comprises a first fixed sequence oligonucleotide 406, 408 comprising sequences complementary to loci 410, 412, label binding regions 414, 416, capture regions 418, 420, universal primer regions 422, 424 and restriction enzyme recognition site regions 426, 428. The label binding regions 414, 416 comprise sequences that are different for the sets of fixed sequence oligonucleotides 402, 404 to allow differential labeling of fixed sequence oligonucleotides associated with each different target genomic region, while the capture regions 418, 420 in this embodiment are the same for both sets of fixed sequence oligonucleotides 402, 404 to allow for competitive hybridization to the capture features of a hybridization array. The restriction enzyme recognition sites 426, 428

can be the same for both sets of fixed sequence oligonucleotides 402, 404 or different for each set depending on the embodiment.

[000119] In step 442, the sets of fixed sequence oligonucleotides 402, 404 are introduced to a sample and allowed to hybridize to elected loci 444, 446. Following hybridization and/or ligation, unhybridized fixed sequence oligonucleotides preferably are separated from the remainder of the sample (not shown). In step 448, the sets of fixed sequence oligonucleotides 402, 404 are ligated to create ligation products 450, 452. In step 454, universal primers 456, 458, 460 and 462 are introduced to the ligation products 450, 452 which bind to universal primer regions 422, 438, 424, and 440, respectively, to create amplicons 464, 466 comprising label binding regions 472, 474, capture regions 468, 470 and restriction enzyme recognition sites 476, 478. In step 480, a restriction enzyme is introduced to the amplicons 464, 466 which binds to the restriction enzyme recognition site 476, 478 and cleaves the amplicons leaving a cleaved amplicon 482, 484 comprising the label binding regions 472, 474 and the capture regions 468, 470. Also at step 480, the cleaved amplicons 482, 484 are introduced to a hybridization array 486 comprising a plurality of capture probes 488 where the capture regions 468, 470 of the cleaved amplicons 482, 484 competitively hybridize to the capture probes 488. In step 492, labeled oligonucleotides 492, 494 are introduced to the array 486 where the label binding regions 472, 474 of each cleaved amplicon 482, 484 hybridize to target recognition regions 496, 497 of the labeled oligonucleotides 492, 494. Following hybridization of the labeled oligonucleotides to the cleaved amplicons, unhybridized labeled oligonucleotides preferably are separated from the array (not shown). The labels 498, 499 of the labeled oligonucleotides 492, 494 can then be detected and the cleaved amplicons corresponding to each target genomic region can be quantified to provide information on the presence and amount of the target genomic regions in the sample. Note that in Fig. 4 labeled oligonucleotides 492, 494 abut capture probes 488, and thus labeled oligonucleotides 492, 494 can be ligated to capture

probes 488 to, *e.g.*, increase binding stability. Cleaved amplicons 482, 484 can then be eliminated from the array by washing.

[000120] FIG. 5 is an illustration of another specific embodiment of the invention in which the ligation product is cleaved prior to hybridization to an array. In the method depicted in FIG. 5, two sets of fixed sequence oligonucleotides 502, 504 are provided. Each set comprises a first fixed sequence oligonucleotide 506, 508 comprising sequences complementary to loci 510, 512, label binding regions 514, 516, capture regions 518, 520, universal primer regions 522, 524 and restriction enzyme recognition site regions 526, 528. The label binding regions 514, 516 comprise sequences that are different for the sets of fixed sequence oligonucleotides 502, 504 to allow differential labeling of fixed sequence oligonucleotides associated with each different target genomic region, while the capture regions 518, 520 in this embodiment are the same for both sets of fixed sequence oligonucleotides 502, 504 to allow for competitive hybridization to the capture features of a hybridization array. The restriction enzyme recognition sites 526, 528 can be the same for both sets of fixed sequence oligonucleotides 502, 504 or different for each set depending on the embodiment.

[000121] In step 542, the sets of fixed sequence oligonucleotides 502, 504 are introduced to a sample and allowed to hybridize loci 544, 546. Following hybridization, unhybridized fixed sequence oligonucleotides preferably are separated from the remainder of the sample (not shown). In step 548, the sets of fixed sequence oligonucleotides 502, 504 are ligated to create ligation products 550, 552. In step 554, universal primers 556, 558, 560 and 562 are introduced to the ligation products 550, 552 which bind to universal primer regions 522, 538, 524, and 540, respectively, to create amplicons 564, 566 comprising label binding regions 572, 574, capture regions 568, 570 and restriction enzyme recognition sites 576, 578. In step 580, a restriction enzyme is introduced to the amplicons 564, 566 which binds to the restriction enzyme recognition site 576, 578 and cleaves the amplicons leaving a cleaved amplicon 582, 584 comprising the label binding regions 572, 574 and the capture regions 568, 570. The cleaved

products are bound 580 to their respective labels 582, 584 and introduced to a hybridization array 586 comprising a plurality of capture probes 588 where the capture regions 568, 570 of the cleaved amplicons competitively hybridize 588 to capture probes 590.

[000122] In certain aspects of this embodiment of the invention (and as illustrated in and discussed in relation to in Fig. 4), the labeled oligonucleotides may be ligated to the capture probes of the hybridization array to increase the binding stability. This embodiment requires juxtaposition of the labelled oligonucleotide and the capture probe. The capture probes, cleaved ligation products and labeled oligonucleotides may be configured so that the labeled oligonucleotides and capture probes are hybridized adjacently to one another, or the labeled oligonucleotides and the capture probes can be extended or a gap-filling oligonucleotide may be employed as described elsewhere herein.

Detection of Copy Number Variations

[000123] As stated above, the invention provides methods to identify copy number variants of target genomic regions (relatively short genomic regions, larger genomic regions, sub-chromosomal regions and chromosomes), mutations, and polymorphisms in a sample. The present invention provides particularly powerful methods for identifying fetal chromosomal aneuploidies in maternal samples comprising both maternal and fetal DNA.

[000124] The methods of the invention can also be used to analyze multiple loci on multiple chromosomes and average the frequency of the loci from a particular chromosome together. Normalization or standardization of the frequencies can be performed for one or more target sequences.

[000125] In some aspects, the methods of the invention can be used to sum the frequencies of the loci on each chromosome for both sources in a mixed sample such as a maternal serum sample, *e.g.*, by detecting an overall signal of a label, and then comparing the sum of the labels from the loci on one chromosome to the sum of the labels from the loci on another chromosome

to determine whether a chromosomal abnormality exists. Alternatively, one can analyze subsets of loci on each chromosome to determine whether a chromosomal abnormality exists. The comparison can be made either between regions from the same chromosome or between regions from different chromosomes.

[000126] The data used to determine the frequency of the loci may exclude outlier data that appear to be due to experimental error, or that have elevated or depressed levels based on an idiopathic genetic bias within a particular sample. In one example, the data used for summation may exclude DNA regions with a particularly elevated frequency in one or more samples. In another example, the data used for summation may exclude loci that are found in a particularly low abundance in one or more samples.

[000127] Subsets of loci can be chosen randomly but with sufficient numbers to yield a statistically significant result in determining whether a chromosomal abnormality exists. Multiple analyses of different subsets of loci can be performed within a mixed sample and on different arrays to yield more statistical power. For example, if there are 100 loci for chromosome 21 and 100 loci for chromosome 18, a series of analyses could be performed that evaluate fewer than 100 loci for each of the chromosomes on different arrays. In this example, target sequences are not being selectively excluded.

[000128] The quantity of different loci detectable on certain chromosomes may vary depending upon a number of factors, including general representation of fetal loci in maternal samples, degradation rates of the different loci representing fetal loci in maternal samples, sample preparation methods, and the like.

Tandem Ligation Assay

[000129] In certain aspects, the methods of the invention employ tandem ligation methods comprising the use of first and second fixed sequence oligonucleotides that are complimentary to loci in a target genomic region, *e.g.,* a chromosome of interest or a reference chromosome,

and one or more short, bridging oligonucleotides (also called “splint” oligonucleotides or “gap” or “gap-filling” oligonucleotides) complementary to the regions of the loci between and immediately adjacent to the first and second fixed sequence oligonucleotides. Hybridization of these oligonucleotides between hybridized fixed sequence oligonucleotides on a selected locus, followed by ligation of these oligonucleotides, provides a ligation product which in turn provides a template for amplification, if desired. A tandem ligation assay tends to be more discriminating than use of a single oligonucleotide probe or use of only two fixed oligonucleotide probes, as perfect complementarity between the fixed sequence and bridging oligonucleotides and the selected locus must exist at both ligation sites for ligation to occur.

[000130] In preferred aspects, a single bridging oligonucleotide is used which hybridizes adjacently between the fixed sequence oligonucleotides in the tandem ligation methods. Alternatively in some aspects, the tandem ligation methods use sets of two fixed sequence oligonucleotides with a set of two or more bridging oligonucleotides that hybridize adjacently in the region of the nucleic acid between the region complementary to the first and second fixed sequence oligonucleotides. These bridging oligonucleotides hybridize adjacent to one another and to the fixed sequence oligonucleotides. The two fixed sequence oligonucleotides and two bridging oligonucleotides are ligated during the ligation reaction, resulting in a single ligation product which serves as a template for amplification and, ultimately for detection.

[000131] In other aspects of the invention, the method employs sets of fixed sequence oligonucleotides that bind to non-adjacent regions within a selected locus, and primer extension is utilized to create a contiguous set of hybridized oligonucleotides prior to the ligation step. Alternatively, the methods may employ sets of fixed sequence oligonucleotides, one or more bridging oligonucleotides and extension of one or more of the hybridized oligonucleotides. The combination of extension and ligation provides a ligation product which in turn serves as a template for amplification, followed by detection and quantification.

[000132] In a preferred aspect, the methods of the invention employ a multiplexed reaction with a set of three or more oligonucleotides for each selected locus. This general aspect is illustrated in FIG. 6 in which two sets of fixed sequence oligonucleotides 602, 604 are provided. Each set of fixed sequence oligonucleotides 602, 604 comprises first fixed sequence oligonucleotides 606, 608 each comprising a sequence complementary to a nucleic acid region of interest 610, 612, a label binding region 614, 616 and a universal primer region 618, 620, and second fixed sequence oligonucleotides 622, 624 each comprising a sequence complementary to the nucleic acid region of interest 626, 628, a capture region 630, 632 and a universal primer region 634, 636. In step 638, the sets of fixed sequence oligonucleotides 602, 604 are introduced to a sample and allowed to specifically hybridize to complementary portions of loci 640, 642 in the target genomic regions. Following hybridization, unhybridized fixed sequence oligonucleotides preferably are separated from the remainder of the genetic sample (not shown). The removal of the unhybridized fixed sequence oligonucleotides may alternatively be separated from the remainder of the genetic sample after the ligation step below. In step 644, sets of bridging oligonucleotides 646, 648 are introduced and allowed to specifically hybridize to the region of the loci between the first fixed sequence oligonucleotide 606, 608 and second fixed sequence oligonucleotide 622, 624. Alternatively, the bridging oligonucleotides 646, 648 can be introduced simultaneously with the fixed sequence oligonucleotides.

[000133] In step 650, the sets of fixed sequence oligonucleotides 602, 604 and bridging oligonucleotides 630, 632 are ligated to create ligation products 652, 654 complementary to the loci 640, 642. Also, in step 650, universal primers 658, 660, 662 and 664 are introduced to the ligation products 652, 654 where the universal primers bind to the universal primer regions 618, 634, 620, and 636, respectively, and amplify the ligation products to create amplicons 678, 680 comprising label binding regions 666, 668 and capture regions 662, 664. In certain preferred embodiments, universal primers 658 and 662 have substantially the same sequence, which is complementary to both 618 and 620, and universal primers 660 and 664 have substantially the

same sequence, which is complementary to both 634 and 636. The amplicons 678, 680 are introduced to a hybridization array 670 comprising a plurality of capture probes 672 wherein the capture regions 662, 664 of the amplicons 658, 660 competitively hybridize to target capture regions 674 on the capture probes 672. In step 676, target recognition regions 682, 684 of the labeled oligonucleotides 688, 690 specifically hybridize to the label binding regions 666, 668 of the amplicons 678, 680. Following hybridization of the labeled oligonucleotides, unhybridized labeled oligonucleotides preferably are removed from the array (not shown). The labels 686, 688 of the labeled oligonucleotides 688, 690 can then be detected and the loci optionally quantified to provide information on the presence and amount of each genomic region in the genetic sample.

[000134] In certain aspects, the bridging oligonucleotides can be composed of a mixture of oligonucleotides with degeneracy in each of the positions, so that the mixture of bridging oligonucleotides will be compatible with all reactions in the multiplexed assay requiring a bridging of a given length. In one example the bridging oligonucleotide is a randomer, where all combinations of the bridging oligonucleotide are synthesized. As an example, in the case where a 5-base oligonucleotide is used, the number of unique bridging oligonucleotide s would be 4^5 = 1024. This would be independent of the number of targeted regions since all possible bridging oligonucleotide s would be present in the reaction. In another aspect, the bridging oligonucleotides can be of various lengths so that the mixture of oligonucleotides is compatible with ligation reactions requiring bridging oligonucleotides of different lengths.

[000135] In yet another aspect, the bridging oligonucleotide can have degeneracy at specific positions—*i.e.*, known polymorphic sites—and the tandem ligation reactions are restricted to those that require a specific sequence provided at the polymorphic site.

[000136] In another example, the bridging oligonucleotide is specific, synthesized to match the sequences in the gap. As an example, in the case where a 5-base oligonucleotide is used, the number of unique oligonucleotides provided in the assay would be equal to or less than the

number of loci. A number of bridging oligonucleotides less than the number of loci could be achieved if the gap sequence was shared between two or more loci. In one aspect of this example, one might purposefully choose loci and especially the gap sequences such that there was as much identical overlap as possible in the gap sequences, minimizing the number of bridging oligonucleotides necessary for the multiplexed reaction.

[000137] In another aspect, the sequences of the bridging oligonucleotides are designed and the loci are selected so that all loci share the same base(s) at each end of the bridging oligonucleotide. For instance, one might choose loci and their gap location such that all of the gaps shared an “A” base at the first position and a “G” base at the last position of the gap. Any combination of a first and last base could be utilized, based upon factors such as the genome investigated, the likelihood of sequence variation in that area, and the like. In a specific aspect of this example, the bridging oligonucleotides can be synthesized by random degeneracy of bases at the internal positions of the bridging oligonucleotide, with nucleotides specific at the first and last position of the bridging oligonucleotide. In the case of a 5-mer, the second, third and fourth positions would be degenerate, and the two specific nucleotides at the end of the bridging oligonucleotide would be fixed. In this case, the number of unique bridging oligonucleotides would be 4^3 = 64.

[000138] In the human genome the frequency of the dinucleotide CG is much lower than expected by the respective mononucleotide frequencies. This presents an opportunity to enhance the specificity of an assay with a particular mixture of bridging oligonucleotides. In this aspect, the bridging oligonucleotides may be selected to have a 5' G and a 3' C. This base selection allows each oligonucleotide to have a high frequency in the human genome but makes it a rare event for two bridging oligonucleotides to hybridize adjacent to each other. The probability is then reduced that multiple oligonucleotides are ligated in locations of the genome that are not targeted in the assay.

[000139] In certain aspects, the bridging oligonucleotide is added to the reaction after the sets of fixed sequence oligonucleotides have been hybridized, and following the optional removal of all unhybridized fixed sequence oligonucleotides. The conditions of the hybridization reaction preferably are optimized near the T_m of the bridging oligonucleotide to prevent erroneous hybridization of oligonucleotides that are not fully complementary to the nucleic acid region. If the bridging oligonucleotides have a T_m significantly lower than the fixed sequence oligonucleotides, the bridging oligonucleotide is preferably added as a part of the ligase reaction.

[000140] The advantage of using short bridging oligonucleotides is that ligation on either end would likely occur only when all bases of the bridging oligonucleotide match the gap sequence. A further advantage of using short bridging oligonucleotides is that the number of different bridging oligonucleotides necessary could be less than the number of loci, raising the oligonucleotides' effective concentration to allow perfect matches to happen faster. Fewer numbers of bridging oligonucleotides provides the advantages of cost savings and quality control. The advantages of using fixed first and last bases and random bases in between include the ability to utilize longer bridging oligonucleotides for greater specificity while reducing the number of total bridging oligonucleotides in the reaction.

Detection of Polymorphisms

[000141] In certain aspects, the methods of the invention detect one or more target genomic regions that comprise a polymorphism. In some embodiments, this methodology is not necessarily designed to identify a particular allele, *e.g.*, as maternal versus fetal, but rather to ensure that different alleles corresponding to loci in a target genomic region are included in the quantification methods of the invention. In certain aspects, however, it may be desirable to both use allelic information to count all loci in the target genomic region as well as to use the allelic information, *e.g.*, to calculate the amount of fetal DNA contained within a maternal sample, or

identify the percent of alleles with a particular mutation in a genetic sample from a cancer patient. Thus, the methods of the invention are intended to encompass both mechanisms for detection of SNP-containing loci for direct determination of copy number variation through quantification as well as detection of SNPs for ensuring overall efficiency of the assay.

[000142] Thus, in a particular aspect of the invention, allelic discrimination is provided through the fixed sequence oligonucleotides or through the bridging oligonucleotides used for detection of the loci. In such embodiments, the label binding region may serve as an allele index that is embedded in either the first fixed sequence oligonucleotide or the second fixed sequence oligonucleotide in a set. In certain specific aspects, an allele index (*e.g.*, label binding region) is present on both the first and second fixed sequence oligonucleotides to detect two or more polymorphisms within the loci. The number of fixed sequence oligonucleotides used in such aspects can correspond to the number of possible alleles being assessed for a selected locus, and detection of a label associated with the allele index can detect presence, amount or absence of a specific allele in a genetic sample.

[000143] FIG. 7 illustrates one aspect of the invention in which polymorphisms are detected using competitive hybridization to an array. In FIG. 7, two sets of fixed sequence oligonucleotides 702, 704 are provided in which the first fixed sequence oligonucleotide 706, 708 of each set comprises a sequence that is complementary to a locus of interest 710, 712 comprising for example an A/T or G/C SNP, respectively and a label 714, 716. The second fixed sequence oligonucleotide 722, 724 comprises a sequence that is complementary to the selected locus 726, 728 and a capture region 730, 732. In some embodiments, fixed sequence oligonucleotide 706 is the same as fixed sequence oligonucleotide 708 except for SNPs 718 and 720 and labels 714 and 716, respectively; and fixed sequence oligonucleotide 722 is the same as fixed sequence oligonucleotide 724. That is, the difference between the sets of fixed sequence oligonucleotides is the SNP interrogated and the label that corresponds to the SNP. In step 734, the sets of fixed sequence oligonucleotides 702, 704 are introduced to the sample

and allowed to specifically hybridize to locus of interest 736, 738 comprising SNPs 740, 742. Following hybridization or ligation, unhybridized fixed sequence oligonucleotides preferably are separated from the remainder of the genetic sample (not shown). In step 744, the sets fixed sequence oligonucleotides 702, 704 are ligated to create ligation products 746, 748. It is important to note that in this embodiment the ligation is allele-specific as long as the allele-specific nucleotide is close to the ligation junction. Typically, the allele-specific nucleotide must be within 5 nucleotides of the ligated end; however, in preferred aspects, the allele-specific nucleotide is the terminal base.

[000144] In step 750, the ligation products 746, 748 are introduced to a hybridization array 752 comprising a plurality of capture probes 754 where the capture regions 730, 732 of the ligation products 746, 748 hybridize to the capture probes 754. Following hybridization of the ligation products, unhybridized ligation products preferably are separated from the remainder of the genetic sample (not shown). The labels 714, 716 can then be detected and alleles specific to the loci corresponding to each target genomic region are quantified to provide information on the presence and amount each allele and target genomic region in the genetic sample.

[000145] In another embodiment, the allele-specific nucleotide is disposed in the first and/or second fixed sequence oligonucleotides and bridging oligonucleotides are used to create ligation products. FIG. 8 illustrates this aspect of the invention. In FIG. 8, two sets of fixed sequence oligonucleotides 802, 804 are provided, where both sets of fixed sequence oligonucleotides are configured to interrogate different SNPs at the same selected locus. The first fixed sequence oligonucleotides 806, 808 of each set comprise a sequence that is complementary to a selected locus 810, 812 comprising for example an A/T or G/C SNP, respectively, a label binding region 814, 816, and a universal primer region 818, 820; and second fixed sequence oligonucleotides 826, 828 each comprising sequences that are complementary to a selected locus 830, 832, capture region 834, 836 and universal primer regions 838, 840. In the polymorphic assays, second fixed sequence oligonucleotides 826 and

828 have substantially the same sequence and first fixed sequence oligonucleotides 806, 808 have substantially the same sequence except for the SNP site and the label binding regions 814, 816. In step 842, the sets of fixed sequence oligonucleotides 802, 804 are introduced to the sample and are allowed to hybridize to the loci 844, 846 comprising SNPs 848, 850. Following hybridization, unhybridized fixed sequence oligonucleotides preferably are separated from the remainder of the genetic sample (not shown).

[000146] In step 852, a set of bridging oligonucleotides 854, 856 are introduced to the sample and allowed to hybridize to the locus 844, 846 between the first and second fixed oligonucleotides of each set. In a preferred embodiment, 854 and 856 have substantially the same sequence. In step 858, the hybridized fixed sequence oligonucleotides and hybridized bridging oligonucleotides are ligated to create ligation products 860, 862. In step 864, universal primers 866, 868, 870 and 872 are introduced to the ligation products 860, 862 which bind to universal primer regions 818, 838, 820, and 840, respectively and amplify the ligation products 860, 862 to produce amplicons 874, 876 comprising label binding regions 882, 884 and capture regions 878, 880. The amplicons are introduced to a hybridization array 886 comprising capture probes 888 where the capture regions 878, 880 of the amplicons 874, 876 hybridize to the capture probes 888. In step 892, labeled oligonucleotides 894, 895 are introduced to the hybridization array 886 under conditions that allow the target recognition regions 896, 897 of the labeled oligonucleotides 894, 895 to selectively hybridize to the label binding regions 882, 884 of the amplicons 874, 876. Following hybridization of the labeled oligonucleotides, unhybridized labeled oligonucleotides preferably are removed from the array (not shown). The labels 898, 899 can then be detected and alleles corresponding to the loci corresponding to the target genomic regions quantified to provide information on the presence and amount of the alleles at loci and the target genomic regions in the genetic sample.

[000147] In certain aspects of the invention, allelic discrimination is provided through the bridging oligonucleotide. In this aspect, the bridging oligonucleotide is located over a SNP,

preferably located close enough to one end of a ligation junction so as to provide allelic-specificity.

[000148] In one example, both allele bridging oligonucleotide variants are present in the same reaction mixture and allele detection results from subsequent hybridization of associated labels to hybridization arrays. FIG. 9 illustrates this aspect.

[000149] In FIG. 9, two sets of fixed sequence oligonucleotides 902, 904 are provided where both sets of fixed sequence oligonucleotides are configured to interrogate the same selected locus and are the same except for the label binding region. In this embodiment, the bridging oligonucleotide interrogates the SNP, and the assay must be performed in two separate vessels at least until after the ligation step has taken place. The first fixed sequence oligonucleotides 906, 908 in each set comprise sequences that are complementary to a locus 910, 912, a capture region 914, 916 and universal primer region 918, 920. The second fixed sequence oligonucleotides 922, 924 of each set comprise sequences that are complementary to the locus 926, 928, a label binding region 930, 932 region and a universal primer region 934, 936. In step 938, the sets of fixed sequence oligonucleotides 902, 904 are introduced to the sample under conditions that allow each set 902, 904 to specifically hybridize to the locus 940, 942 wherein the selected locus comprises a SNP 944, 946. Following hybridization (or alternatively following ligation) of the first and second fixed sequence oligonucleotides, unhybridized fixed sequence oligonucleotides preferably are separated from the remainder of the genetic sample (not shown). In step 948, bridging oligonucleotides 950, 952 corresponding to an A/T SNP or a G/C SNP are introduced and allowed to bind to the region of the locus 940, 942 between the first and second fixed sequence oligonucleotides. Alternatively, the bridging oligonucleotides can be introduced to the sample simultaneously with the fixed sequence oligonucleotides.

[000150] In step 954, the hybridized fixed sequence oligonucleotides and bridging oligonucleotides are ligated to create ligation products 956, 958. At this point the two separate

assays optionally can be combined into a single vessel but need not be. In step 960, universal amplification primers 962, 964, 966 and 968 are introduced to the ligation products 956, 958 where the universal primers bind to universal primer regions 918, 934, 920, and 936, respectively, and amplify the ligation products 956, 958 to produce amplicons 970, 972, each comprising a label binding region 980, 982 and a capture region 974, 976. The amplicons 970, 972 are introduced to a hybridization array 984 comprising a plurality of capture probes 986 wherein the capture regions 974, 976 of the amplicons 970, 972 competitively hybridize to the capture probes 986. In step 990, labeled oligonucleotides 992, 994 are introduced to the array 984 where target recognition regions 996, 997 of the labeled oligonucleotides 992, 994 hybridize to label binding regions 980, 982 of the amplicons 970, 972. Following hybridization of the labeled oligonucleotides, unhybridized labeled oligonucleotides preferably are removed from the array (not shown). The labels 998, 999 can then be detected and alleles corresponding to the loci corresponding in turn to the target genomic regions are quantified to provide information on the presence and amount of each allele and target genomic region in the sample.

Additional Embodiments

[000151] FIG. 10 is an illustration of another specific embodiment of the invention in which bridging oligonucleotides are used and the ligation products are dually cleaved prior to hybridization to an array. In the method depicted in FIG. 10, two sets of fixed sequence oligonucleotides 1002, 1004 are provided. Each set comprises a first fixed sequence oligonucleotide 1006, 1008 comprising sequences complementary to loci 1010, 1012, label binding regions 1018, 1020, capture regions 1014, 1016, universal primer regions 1022, 1024 and restriction enzyme recognition site regions 1026, 1028. The label binding regions 1018, 1020 comprise sequences that are different for the sets of fixed sequence oligonucleotides 1002, 1004 to allow differential labeling of fixed sequence oligonucleotides associated with each different target genomic region, while the capture regions 1014, 1016 in this embodiment are

the same for both sets of fixed sequence oligonucleotides 1002, 1004 to allow for competitive hybridization to the capture features of a hybridization array. The restriction enzyme recognition sites 1026, 1028 can be the same for both cleavage sites on each fixed sequence oligonucleotide, and/or the same for the different sets of fixed sequence oligonucleotides 1002, 1004 depending on the embodiment. There is also a second fixed sequence oligonucleotide 1030, 1032 in each set, where each second fixed sequence oligonucleotide comprises a sequence complementary to the loci 1034, 1036 and a universal primer region 1038, 1040.

[000152] In step 1042, the sets of fixed sequence oligonucleotides 1002, 1004 are introduced to a sample and allowed to hybridize loci 1044, 1046. Following hybridization (or alternatively ligation), unhybridized fixed sequence oligonucleotides preferably are separated from the remainder of the sample (not shown). In step 1048, a bridging oligonucleotide 1026, 1028 is added to each set and hybridized adjacently between the hybridized fixed sequence oligonucleotides. At step 1054, the sets of oligonucleotides are ligated to create ligation products 1050, 1052. In step 1054, universal primers 1056, 1058, 1060 and 1062 are introduced to the ligation products 1050, 1052 which bind to universal primer regions 1022, 1038, 1024, and 1040, respectively, to create amplicons 1064, 1066 comprising label binding regions 1018, 1020, capture regions 1014, 1016, and restriction enzyme recognition sites 1072, 1074, 1076, and 1078. In step 1070, one or more restriction enzymes are introduced to the amplicons 1064, 1066 and the amplicons are dually cleaved to create a cleaved amplicon comprising the label binding regions 1018, 1020 and the capture regions 1014, 1016. The cleaved products are introduced to a hybridization array 1086 comprising a plurality of capture probes 1088 where the capture regions 1014, 1016 of the cleaved amplicons competitively hybridize to capture probes 1090. The hybridized capture regions are then introduced 1088 to the labeled oligonucleotides 1082, 1084, which bind to complementary sequences on the cleaved products and are detected.

[000153] FIG. 11 is an illustration of another specific embodiment of the invention in which bridging oligonucleotides are used to identify different alleles in the same locus and the ligation products are dually cleaved prior to hybridization to an array. In the method depicted in FIG. 11, two sets of fixed sequence oligonucleotides 1102, 1104 are provided for a single locus having a possible polymorphism. Each set comprises a first fixed sequence oligonucleotide 1106, 1108 comprising sequences complementary to one allele of the selected locus 1110, 1112, label binding regions 1118, 1120, capture regions 1114, 1116, universal primer regions 1122, 1124 and restriction enzyme recognition site regions 1126, 1128. The label binding regions 1118, 1120 comprise sequences that are different for the sets of fixed sequence oligonucleotides 1102, 1104 to allow differential labeling of fixed sequence oligonucleotides associated with each different allele, while the capture regions 1114, 1116 in this embodiment are the same for both sets of fixed sequence oligonucleotides 1102, 1104 to allow for competitive hybridization to the capture features of a hybridization array. The restriction enzyme recognition sites 1126, 1128 can be the same for both cleavage sites on each fixed sequence oligonucleotide, and/or the same for the different sets of fixed sequence oligonucleotides 1102, 1104 depending on the embodiment. There is also a second fixed sequence oligonucleotide 1130, 1132 in each set, where each second fixed sequence oligonucleotide comprises a sequence complementary to the selected locus 1134, 1136 and a universal primer region 1138, 1140.

[000154] In step 1142, the sets of fixed sequence oligonucleotides 1102, 1104 are introduced to a sample and allowed to hybridize to the selected locus 1144, 1146. Following hybridization (or alternatively ligation), unhybridized fixed sequence oligonucleotides preferably are separated from the remainder of the sample (not shown). In step 1148, a bridging oligonucleotide 1126, 1128 is added to each set and hybridized adjacently between the hybridized fixed sequence oligonucleotides. At step 1154, the sets of oligonucleotides are ligated to create ligation products 1150, 1152. In step 1154, universal primers 1156, 1158, 1160 and 1162 are introduced to the ligation products 1150, 1152 which bind to universal primer

regions 1122, 1138, 1124, and 1140, respectively, to create amplicons 1164, 1166 comprising label binding regions 1118, 1120, capture regions 1114, 1116, and restriction enzyme recognition sites 1172, 1174, 1176, and 1178. In step 1170, one or more restriction enzymes are introduced to the amplicons 1164, 1166 and the amplicons dually cleaved to create a cleaved amplicon comprising the label binding regions 1118, 1120 and the capture regions 1114, 1116. The cleaved products are introduced to a hybridization array 1186 comprising a plurality of capture probes 1188 where the capture regions 1118, 1120 of the cleaved amplicons competitively hybridize to capture probes 1190. The hybridized capture regions are then introduced 1188 to the labeled oligonucleotides 1182, 1184, which bind to complementary sequences on the cleaved products and are detected to differentiate between the different alleles of the locus.

[000155] FIG. 12 is an illustration of another specific embodiment of the invention in which bridging oligonucleotides are used and the ligation products are cleaved prior to hybridization to an array. In this embodiment, the primers used to amplify the ligation products are differentially labeled. In the method depicted in FIG. 12, two sets of fixed sequence oligonucleotides 1202, 1204 are provided. Each set comprises a first fixed sequence oligonucleotide 1206, 1208 comprising sequences complementary to loci 1210, 1212, primer binding regions 1218, 1224 (which are different), capture regions 1214, 1220, and restriction enzyme recognition site regions 1226, 1228. The primer binding regions 1218, 1224 comprise sequences that are different for the sets of fixed sequence oligonucleotides 1202, 1204 to allow differential labeling of fixed sequence oligonucleotides associated with each different target genomic region, while the capture regions 1214, 1220 in this embodiment are the same for both sets of fixed sequence oligonucleotides 1202, 1204 to allow for competitive hybridization to the capture features of a hybridization array. The restriction enzyme recognition sites 1226, 1228 can be the same for both cleavage sites on each fixed sequence oligonucleotide, and/or the same for the different sets of fixed sequence oligonucleotides 1302, 1304 depending on the embodiment. There is

also a second fixed sequence oligonucleotide 1230, 1232 in each set, where each second fixed sequence oligonucleotide comprises a sequence complementary to the loci 1234, 1236 and a primer region 1238, 1240.

[000156] In step 1242, the sets of fixed sequence oligonucleotides 1202, 1204 are introduced to a sample and allowed to hybridize loci 1244, 1246. Following hybridization (or alternatively ligation), unhybridized fixed sequence oligonucleotides preferably are separated from the remainder of the sample (not shown). In step 1248, a bridging oligonucleotide 1226, 1228 is added to each set and hybridized adjacently between the hybridized fixed sequence oligonucleotides. At step 1254, the sets of oligonucleotides are ligated to create ligation products 1250, 1252. In step 1254, primers 1256, 1258, 1260 and 1262 are introduced to the ligation products 1250, 1252 which bind to primer regions 1218, 1238, 1224, and 1240, respectively, to create amplicons 1264, 1266 comprising primer binding regions 1278, 1284, capture regions 1276, 1280, and restriction enzyme recognition sites 1272 and 1274. Primers 1256 and 1260 are differentially labeled thus allowing for differentiation between amplicons from the different loci once the amplicons are hybridized to an array. In step 1270, one or more restriction enzymes are introduced to the amplicons 1264, 1266 and the amplicons are cleaved to create a cleaved amplicon comprising the primer binding regions 1278, 1284 and the capture regions 1276, 1280. The cleaved products are introduced to a hybridization array 1286 at step 1290 comprising a plurality of capture probes 1288 where the capture regions 1276, 1280 of the cleaved amplicons competitively hybridize to capture probes 1288.

[000157] FIG. 13 is an illustration of another specific embodiment of the invention in which bridging oligonucleotides are used and the ligation products are cleaved prior to hybridization to an array. FIG 13 illustrates an embodiment very similar to that in Fig. 12. In this embodiment, the primers used to amplify the ligation products are differentially labeled. In the method depicted in FIG. 13, two sets of fixed sequence oligonucleotides 1302, 1304 are provided. Each set comprises a first fixed sequence oligonucleotide 1306, 1308 comprising sequences

complementary to loci 1310, 1312, primer binding regions 1318, 1324 (which are different), capture regions 1314, 1320 (which are the same), and restriction enzyme recognition site regions 1326, 1328. The primer binding regions 1318, 1324 comprise sequences that are different for the sets of fixed sequence oligonucleotides 1302, 1304 to allow differential labeling of fixed sequence oligonucleotides associated with each different target genomic region, while the capture regions 1314, 1320 in this embodiment are the same for both sets of fixed sequence oligonucleotides 1302, 1304 to allow for competitive hybridization to the capture features of a hybridization array. The restriction enzyme recognition sites 1326, 1328 can be the same for both cleavage sites on each fixed sequence oligonucleotide, and/or the same for the different sets of fixed sequence oligonucleotides 1302, 1304 depending on the embodiment. There is also a second fixed sequence oligonucleotide 1330, 1332 in each set, where each second fixed sequence oligonucleotide comprises a sequence complementary to the loci 1334, 1326 and a primer region 1338, 1340.

[000158] In step 1342, the sets of fixed sequence oligonucleotides 1302, 1304 are introduced to a sample and allowed to hybridize loci 1344, 1346. Following hybridization (or alternatively ligation), unhybridized fixed sequence oligonucleotides preferably are separated from the remainder of the sample (not shown). In step 1348, a bridging oligonucleotide 1326, 1328 is added to each set and hybridized adjacently between the hybridized fixed sequence oligonucleotides. At step 1354, the sets of oligonucleotides are ligated to create ligation products 1350, 1352. In step 1354, primers 1356, 1358, 1360 and 1362 are introduced to the ligation products 1350, 1352 which bind to primer regions 1318, 1338, 1324, and 1340, respectively, to create amplicons 1364, 1366 comprising primer binding regions 1378, 1384, capture regions 1376, 1380, and restriction enzyme recognition sites 1372 and 1374. Primers 1356 and 1360 are differentially labeled thus allowing for differentiation between amplicons from the different loci once the amplicons are hybridized to an array. In step 1370, one or more restriction enzymes are introduced to the amplicons 1364, 1366 and the amplicons are cleaved

to create a cleaved amplicon comprising the primer binding regions 1378, 1384 and the capture regions 1376, 1380. The cleaved products are introduced to a hybridization array 1386 at step 1390 comprising a plurality of capture probes 1388 where the capture regions 1376, 1380 of the cleaved amplicons competitively hybridize to capture probes 1388.

[000159] In the embodiments of FIGs. 10-13, the cleavage amplicons introduced to the array do not include any portion of the target genomic region or a sequence complementary thereof. The capture regions are used in association with different target genomic regions, and the labeled oligonucleotide is used to differentiate between regions. In this manner, the cleaved amplicons competitively hybridize to common capture probes, and the quantitation of the target genomic regions is determined by the detected labels corresponding to cleaved products that bind to the array.

[000160] As described previously, the embodiments of the present invention shown in Figures 2-13 depict two separate fixed sequence oligonucleotides used to interrogate each locus or allele. However, in some aspects, however, a single probe can be used which comprises two or more distinct non-adjacent fixed sequence oligonucleotides that are complementary to the loci including precircle probes. Such precircle probes are described, *e.g.,* by Lizardi in US Pat Nos. 5,854,033 and 6,316,229, and can be linearized prior to hybridization to the array by, e.g., including a site for a restriction endonuclease in the probe.

Universal Amplification

[000161] In certain aspects of the invention, universal amplification is used to amplify the ligation products following hybridization and ligation of the fixed sequence oligonucleotides, either directly or following extension or the introduction of a bridging oligonucleotide. In a multiplexed assay system, amplification preferably is done through universal amplification using universal primers that hybridize to universal primer regions on the first and second fixed sequence oligonucleotide of each set. The universal primer regions from the fixed sequence

oligonucleotides become a part of the ligation products, where the ligation products can then be amplified in a single universal amplification reaction. Although these universal primer sequences preferably are introduced via the fixed sequence oligonucleotides, they may also be added to the proximal ends of the ligation products via ligation. The introduction of universal primer regions to the fixed sequence oligonucleotides allows a subsequent controlled universal amplification of all or a portion of the ligation products prior to array hybridization. The amplicon produced from this amplification process can be used directly or further processed prior to introduction to the array for detection. In specific embodiments, the amplicon is cleaved, and the portion comprising the capture region is introduced to the array for hybridization and detection. Bias and variability can be introduced during DNA amplification, such as that seen during polymerase chain reaction (PCR). In cases where an amplification reaction is multiplexed, there is the potential that loci will amplify at different rates or efficiency. Sets of primers for a given locus may behave differently based on sequence context of the primer and template DNA, buffer conditions, and other conditions. In certain aspects, the universal primer regions used in the methods are designed to be compatible with conventional multiplexed assay methods that utilize general priming mechanisms to analyze large numbers of nucleic acids simultaneously. Such “universal” priming methods allow for efficient, high volume analysis of the quantity of nucleic acid regions present in a genetic sample, and allow for comprehensive quantification of the presence of nucleic acid regions within such a sample.

[000162] The entirety of a ligation reaction or an aliquot of the ligation reaction may be used for universal amplification. Using an aliquot allows parallel amplification reactions to be undertaken using the same or different conditions (*e.g.,* polymerase, buffers, and the like), *e.g.,* to ensure that bias is not inadvertently introduced due to experimental conditions. In addition, variations in primer concentrations may be used to effectively limit the number of sequence specific amplification cycles. Examples of multiplexing assay methods include, but are not limited to, those described in Oliphant et al., US Pat. No. 7,582,420.

[000163] As described in detail herein, many methods of the invention utilize coupled reactions for multiplex detection where oligonucleotides from an early phase of the multi-step process contain sequences that may be used in a later phase of the process. Processes known in the art for amplifying and/or detecting nucleic acids in samples can be used, alone or in combination, including but not limited to the methods described below. In certain aspects, the methods of the invention utilize one of the following combined selective and universal amplification techniques: (1) LDR coupled to PCR; (2) primary PCR coupled to secondary PCR coupled to LDR; and (3) primary PCR coupled to secondary PCR. Each of these combinations of techniques can utilize probe regions from an early phase in the process that may be used as a primer sequence in a later phase of the process.

[000164] Barany et al., US Pat Nos. 6,852,487, 6,797,470, 6,576,453, 6,534,293, 6,506,594, 6,312,892, 6,268,148, 6,054,564, 6,027,889, 5,830,711, 5,494,810, describe the use of the ligase chain reaction (LCR) assay for the detection of specific sequences of nucleotides in a variety of nucleic acid samples.

[000165] Barany et al., US Pat Nos. 7,807,431, 7,455,965, 7,429,453, 7,364,858, 7,358,048, 7,332,285, 7,320,865, 7,312,039, 7,244,831, 7,198,894, 7,166,434, 7,097,980, 7,083,917, 7,014,994, 6,949,370, 6,852,487, 6,797,470, 6,576,453, 6,534,293, 6,506,594, 6,312,892, and 6,268,148 describe the use of the ligase detection reaction with detection reaction ("LDR") coupled with polymerase chain reaction ("PCR") for nucleic acid detection.

[000166] Barany et al., US Pat No. 7,556,924 and 6,858,412, describe the use of padlock probes (also called “precircle probes” or “multi-inversion probes”) with coupled ligase detection reaction ("LDR") and polymerase chain reaction ("PCR") for nucleic acid detection.

[000167] Barany et al., US Pat Nos. 7,807,431, 7,709,201, and 7,198, 814 describe the use of combined endonuclease cleavage and ligation reactions for the detection of nucleic acid sequences.

[000168] Willis et al., US Pat Nos. 7,700,323 and 6,858,412, describe the use of precircle probes in multiplexed nucleic acid amplification, detection and genotyping.

[000169] Ronaghi et al., US Pat. No. 7,622,281 describes amplification techniques for labeling and amplifying a nucleic acid using an adapter comprising a unique primer and a barcode.

[000170] The nucleic acid regions of interest are identified using hybridization techniques and arrays. Methods for conducting polynucleotide hybridization assays for detection are well-developed and well-known in the art. Hybridization assay procedures and conditions will vary depending on the application and are selected in accordance with the general binding methods known including those referred to in: Maniatis et al. Molecular Cloning: A Laboratory Manual (2nd Ed. Cold Spring Harbor, N.Y., 1989); Berger and Kimmel Methods in Enzymology, Vol. 152, Guide to Molecular Cloning Techniques (Academic Press, Inc., San Diego, Calif., 1987); Young and Davis, P.N.A.S, 80: 1194 (1983). Methods and apparatus for carrying out repeated and controlled hybridization reactions have been described in U.S. Pat. Nos. 5,871,928, 5,874,219, 6,045,996 and 6,386,749, 6,391,623 each of which are incorporated herein by reference.

[000171] In some embodiments, the arrays comprise multiple substrates in solution, such as those taught, *e.g.,* in U.S. Appln No. 20140057269 and U.S. Appln No. 20140042366 and U.S. Appln No. 20140024550.

[000172] The present invention also contemplates signal detection of hybridization between ligands in certain preferred aspects. See U.S. Pat. Nos. 5,143,854, 5,578,832; 5,631,734; 5,834,758; 5,936,324; 5,981,956; 6,025,601; 6,141,096; 6,185,030; 6,201,639; 6,218,803; and 6,225,625, in U.S. Patent application 60/364,731 and in PCT Application PCT/US99/06097 (published as WO99/47964).

[000173] Methods and apparatus for signal detection and processing of intensity data are disclosed in, for example, U.S. Pat. Nos. 5,143,854, 5,547,839, 5,578,832, 5,631,734, 5,800,992, 5,834,758; 5,856,092, 5,902,723, 5,936,324, 5,981,956, 6,025,601, 6,090,555,

6,141,096, 6,185,030, 6,201,639; 6,218,803; and 6,225,625, in U.S. Patent application 60/364,731 and in PCT Application PCT/US99/06097 (published as WO99/47964),.

[000174] In certain aspects, the capture region of the ligation products from a single sample or the amplicons thereof contain index sequences that identify the ligation products as being from a particular sample. The features of the array will comprise capture probes that include sequences complementary to the index sequences of the different samples to allow identification of the loci from a particular sample from which the loci originated.

Estimation of Fetal DNA Proportion in a Maternal Sample

[000175] The proportion of fetal DNA in a maternal sample may be used as a part of the risk calculation of the present invention, as fetal proportion provides important information on the expected statistical presence of chromosomal dosage. Variation from the expected statistical presence may be indicative of fetal aneuploidy, an in particular a fetal trisomy or monosomy of a particular chromosome.

[000176] Any methods known in the art to estimate the percentage of fetal DNA in a maternal sample may be used, including quantifying Y sequences if the fetus is male or looking at epigenetic markers (*see, e.g.,* Chim, et al., PNAS USA, 102:14753-58 (2005)). Using fetal proportion as one component of the risk calculation is particularly helpful in circumstances where the level of fetal DNA in a maternal sample is low. Further, knowledge of the fetal DNA percentage may be used to determine what if any additional analyses can be performed on the sample, as it may be the case at a certain lower bound of fetal DNA percentage a system is not able to reliably perform analysis. In other aspects, determining the fetal DNA proportion in a maternal sample may additionally affect the level of certainty or power in detecting a fetal aneuploidy.

[000177] Although the following methods are described for determination of a total proportion of fetal content in a maternal sample, the proportion can also be determined on a

chromosome by chromosome basis. For instance frequency information for fetal chromosome 21 can be determined as compared to fetal chromosome 18. In another example, two or more chromosomes can be used in detecting a fetal proportion, *e.g.,* frequency of loci on chromosomes 1 and 2 can be used. In certain aspects, the chromosome used for determining fetal proportion is the chromosome interrogated for possible aneuploidy. In another aspect, the chromosome(s) used for determining fetal proportion are specifically not the chromosome interrogated for possible aneuploidy.

[000178] The DNA from a fetus will have approximately 50% of its loci inherited from the mother and approximately 50% its loci inherited from the father. Determining which genetic loci are contributed to the fetus from non-maternal sources (informative loci) allows the estimation of fetal DNA proportion in a maternal sample, and thus provides information used to calculate statistically significant differences in chromosomal dosages for chromosomes of interest.

[000179] In certain aspects, determination of non-maternal polymorphisms is achieved through targeted SNP and/or mutation analysis to estimate the percentage of fetal DNA in a maternal sample—a process which is particularly adaptable to the array analysis of the present invention. The percent fetal cell free DNA in a maternal sample can be quantified using multiplexed SNP detection without prior knowledge of the maternal or paternal genotype. In this aspect, two or more selected polymorphic nucleic acid regions with a known SNP in each region are used. In a preferred aspect, the selected polymorphic nucleic acid regions are located on an autosomal chromosome that is unlikely to be aneuploid, *e.g.*, not chromosomes 21, 18, or 13. The selected polymorphic nucleic acid regions from the maternal sample (*e.g.*, plasma) are amplified. In a preferred aspect, the amplification is universal; and in a preferred embodiment, the selected polymorphic nucleic acid regions are amplified in one reaction in one vessel, and in a preferred embodiment, the selected polymorphic nucleic acid regions are amplified in one reaction in one vessel with the ligation products of the fixed sequence oligonucleotides used to determine a chromosomal aneuploidy. Each allele of the selected polymorphic nucleic acid

regions in the maternal sample is determined and quantified. In a preferred aspect, high throughput sequencing is used for such determination and quantification.

[000180] Loci are thus identified where the maternal and non-maternal genotypes are different; *e.g.,* the maternal genotype is homozygous and the non-maternal genotype is heterozygous. This identification of informative loci is accomplished by observing a high frequency of one allele (>80%) and a low frequency (<20% and >0.15%) of the other allele for a particular selected nucleic acid region. The use of multiple loci is particularly advantageous as it reduces the amount of variation in the measurement of the abundance of the alleles between loci. All or a subset of the loci that meet this requirement are used to determine fetal contribution through statistical analysis. In one aspect, fetal contribution is determined by summing the low frequency alleles from two or more loci together, dividing by the sum of the low and high frequency alleles and multiplying by two.

[000181] For many alleles, maternal and non-maternal sequences may be homozygous and identical, and as this information therefore does not distinguish between maternal and non-maternal DNA it is not useful in the determination of percent fetal DNA in a maternal sample. The present invention utilizes allelic information where there is a distinguishable difference between the non-maternal and maternal DNA (*e.g.,* a fetal allele containing at least one allele that differs from the maternal allele) in calculations of percent fetal DNA. Data pertaining to allelic regions that are the same for maternal and non-maternal DNA are thus not selected for analysis, or are removed from the pertinent data prior to estimation of the fetal DNA proportion so as not to mask the useful data. Additional exemplary processes for quantifying fetal DNA in maternal plasma can be found, *e.g.,* in Chu, et al., Prenat. Diagn., 30:1226-29 (2010), which is incorporated herein by reference.

Data Analysis

[000182] Once percent fetal cell free DNA has been calculated, this data may be combined with methods for aneuploidy detection to determine the likelihood that a fetus may contain an aneuploidy. In one aspect, an aneuploidy detection method that utilizes analysis of random DNA segments is used, such as that described in, e.g., Quake, USSN 11/701,686; and Shoemaker et al., USSN No. 12/230,628. In a preferred aspect, aneuploidy detection methods that utilize analysis of selected nucleic acid regions are used. In this aspect, the percent fetal cell free DNA for a sample is calculated. The chromosomal ratio for that sample, a chromosomal ratio for the normal population and a variation for the chromosomal ratio for the normal population is determined, as described herein. Alternatively, the calculated chromosomal ration uses an expectation for a chromosomally normal sample and an expectation for an aneuploid sample. The calculated chromosomal ratio for a sample is then compared to those expectations.

[000183] In one preferred aspect, the chromosomal ratio and its variation for the normal population are determined from normal samples that have a similar percentage of fetal DNA. An expected aneuploid chromosomal ratio for a DNA sample with that percent fetal cell free DNA is calculated by adding the percent contribution from the aneuploid chromosome. The chromosomal ratio for the sample may then be compared to the chromosomal ratio for the normal population and to the expected aneuploid chromosomal ratio to determine statistically, using the variation of the chromosomal ratio, if the sample is more likely normal or aneuploid, and the statistical probability that it is one or the other.

[000184] In a preferred aspect, the selected regions of a maternal sample include both regions for estimation of fetal DNA content as well as non-polymorphic regions from two or more chromosomes to detect a fetal aneuploidy in a single reaction. The single reaction helps to minimize the risk of contamination or bias that may be introduced during various steps in the assay system which may otherwise skew results when utilizing fetal DNA content to help determine the presence or absence of a chromosomal abnormality.

[000185] In other aspects, a selected nucleic acid region or regions may be utilized both for estimation of fetal DNA content as well as detection of fetal chromosomal abnormalities. The alleles for selected nucleic acid regions can be used to estimate fetal DNA content and these same selected nucleic acid regions can then be used to detect fetal chromosomal abnormalities ignoring the allelic information. Utilizing the same selected nucleic acid regions for both fetal DNA content and detection of chromosomal abnormalities may further help minimize any bias due to experimental error or contamination.

[000186] In one embodiment, fetal source contribution in a maternal sample regardless of fetal gender is measured using autosomal SNPs (see, Sparks, et al., Am. J. Obstet & Gyn., 206:319.e1-9 (2012)). The processes utilized do not require prior knowledge of paternal genotype, as the non-maternal alleles are identified during the methods without regard to knowledge of paternal inheritance. A maximum likelihood estimate using the binomial distribution may be used to calculate the estimated fetal nucleic acid contribution across several informative loci in each maternal sample. The processes for calculation of fetal acid contribution used are described, for example, in US Pub. No. 2013/0024127. The polymorphic regions used for determination of fetal contribution may be from chromosomes 1-12, and preferably do not target the blood group antigens. The estimate of fetal contribution from the polymorphic assays is used to define expected response magnitudes when a test chromosome is trisomic, which informs the statistical testing. The test statistic may consist of two components: a measure of deviation from the expected proportion when the sample is disomic; and a measure of deviation from the expected proportion when the sample is trisomic. Each component is in the form of a Wald statistic (e.g., Harrell, Regression modeling strategies, (2001, Springer-Verlag), Sections 9.2.2 and 10.5) which compares an observed proportion to an expected proportion and divides by the variation of the observation.

[000187] The statistic Wj may be used to measure the deviation from expectation when the sample j is disomic, and is defined as

$$W_j = \frac{p_j - p_0}{\sigma_{p_j}},$$

where pj and p0 are defined as described supra with the Z statistic, and σ_{p_j} is the standard deviation of the observed proportion of representation for a given chromosome of interest. The standard deviation may be estimated using parametric bootstrap sampling to create a distribution of pj proportions based on the mean counts and standard errors for our chromosomes of interest. The second statistic is $\widehat{W}_j$, which replaces p0 with the fetal fraction adjusted reference proportion $\hat{p}_j$ is defined as

$$\hat{p}_j = \frac{(1 + 0.5f_j)p_0}{((1 + 0.5f_j)p_0)(1 - p_0)},$$

where f_j is the fetal fraction for sample j and p_0 is the reference proportion as before. This adjustment accounts for the increased representation of a test chromosome when the fetus was trisomic. Because this variance of counts across many loci is measured as a natural result of using multiple non-polymorphic assays for ther test chromosomes, all estimates are taken within a nascent data set and do not require external reference samples or historical information with normalizing adjustments to control for process drift as is typically required for variance around the expected proportion.

[000188] The final statistic used was $S_j = W_j + \widehat{W}_j$. Conceptually, deviations from disomic expectation and trisomic expectation are simultaneously evaluated and summarized into this single statistic. The particular advantage of combining these two indicators is that while deviation from disomy might be high, it may not reach the deviation expected for trisomy at a particular fetal contribution level. The $\widehat{W}_j$ component will be negative in this case, in effect penalizing the deviation from disomy. An $S_j = 0$ indicated an equal chance of being disomic vs. trisomic.

Computer Implementation of the Processes of the Invention

[000189] According to an exemplary embodiment, a computer executes a software component that calculates fetal proportion and applies this information to the values of the dosage of genomic regions and/or chromosomes. In one embodiment, the computer may comprise a personal computer, but the computer may comprise any type of machine that includes at least one processor and memory.

[000190] The output of the software component comprises a report with a value of probability that a genomic region and/or a chromosome has a dosage abnormality. In a preferred aspect this report is an odds ratio of a value of the likelihood that a region or chromosome has two copies (e.g., is disomic) and a value of the likelihood that a region or chromosome has more copies (e.g., is trisomic) or less copies (e.g., is monosomic) copies. The report may be paper that is printed out, or electronic, which may be displayed on a monitor and/or communicated electronically to users via e-mail, FTP, text messaging, posted on a server, and the like.

[000191] Although the normalization process of the invention is shown as being implemented as software, it can also be implemented as a combination of hardware and software. In addition, the software for normalization may be implemented as multiple components operating on the same or different computers.

[000192] Both the server and the computer may include hardware components of typical computing devices, including a processor, input devices (e.g., keyboard, pointing device, microphone for voice commands, buttons, touchscreen, etc.), and output devices (e.g., a display device, speakers, and the like). The server and computer may include computer-readable media, e.g., memory and storage devices (e.g., flash memory, hard drive, optical disk drive, magnetic disk drive, and the like) containing computer instructions that implement the functionality disclosed when executed by the processor. The server and the computer may further include wired or wireless network communication interfaces for communication.

EXAMPLES

[000193] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the present invention, and are not intended to limit the scope of what the inventors regard as their invention, nor are they intended to represent or imply that the experiments below are all of or the only experiments performed. It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific aspects without departing from the spirit or scope of the invention as broadly described. The present aspects are, therefore, to be considered in all respects as illustrative and not restrictive.

[000194] Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees centigrade, and pressure is at or near atmospheric.

Example 1: Subjects

[000195] A total of 878 maternal venous blood samples were analyzed with the following classification of trisomy status: 691 were disomic, 18 were trisomy 13, 37 were trisomy 18, and 132 were trisomy 21. The blood samples were obtained from pregnant women at least 18 years old with singleton pregnancies at 10–34 weeks' gestation. The trisomy classification had previously been determined for all samples tested; 486 samples were originally tested using the Harmony Prenatal Test from Ariosa Diagnostics, Inc. (San Jose, CA), and 396 samples were sourced from patients who underwent either amniotic diagnostic karyotyping or postnatal newborn examination followed by karyotyping when trisomy was suspected.

Example 2: Sample Preparation

[000196] Sample preparation and analysis was performed as described in Sparks, et al., Am J. Obstet Gynecol., 207(5):374.e1-6 (2012). Cell-free DNA was purified from the plasma of

each patient and DANSR™ (Digital Analysis of Selected Regions) assay products (e.g., tandem ligation products) were made from loci on chromosomes 13, 18, and 21. For this analysis, ligation assays using sets of fixed sequence oligonucleotides and bridging oligonucleotides corresponding to 864 genomic regions on each of chromosomes 13, 18, and 21 were used. The DNA sample was attached to a solid support and unhybridized oligonucleotides were removed prior to ligation. In addition, ligation assays using sets of fixed sequence oligonucleotides and bridging oligonucleotides corresponding to 576 polymorphic sites on chromosomes 1 through 12 were developed to evaluate the fraction of fetal cfDNA in each sample. A portion of the ligation assay product produced from each sample was amplified using universal primers and sequenced and the remaining portion of the ligation assay product was hybridized to a custom manufactured DNA array. Prior to hybridization, the amplicons were cleaved. These portions contained identical products.

Example 3: Ligation Product Quantification using Arrays

[000197] Custom DNA arrays were manufactured by Affymetrix, Inc. (Santa Clara, CA) to specifically quantify products of the DANSR assay. DNA arrays were imaged on an Affymetrix GeneTitan® Multi-Channel (MC) Instrument. Each patient sample was assayed on a single custom DNA array. DNA arrays were manufactured and processed in interconnected sets of 384. Next generation sequencing data was produced on an Illumina HiSeq® 2500 (San Diego, CA). Clusters were generated on an Illumina Cluster Station using TruSeq™ Cluster Generation reagents (San Diego, CA). On average 1104 sequencing counts per assay were obtained.

Example 4: Data Analysis

[000198] A previously published algorithm entitled Fetal-Fraction Optimized Risk of Trisomy Evaluation (FORTE™), was used to assign risk scores (see, e.g., Sparks, et al., Am J. Obstet Gynecol., 207(5):374.e1-6(2012)). Non-polymorphic ligation assays on chromosomes

13, 18, and 21 were used to determine chromosome proportion for each of these chromosomes. Polymorphic ligation assays were used to ascertain fetal fraction. Assay variability was defined as the coefficient of variation (CV) of sequence counts (sequencing) or intensities (DNA arrays) of an assay across samples; lower assay variability is preferred. Fetal fraction variability is defined as the relative standard-error of the measured fetal fraction.

Example 5: Results

[000199] For the 878 plasma samples assayed for trisomy risk there was complete concordance between array-based risk scores and trisomy status using the ligation assays and detection by hybridization. In contrast, the concordance between sequence-based risk scores and trisomy status was 99.6%. Sequencing reported high risk scores for two samples that had received low risk scores in a previous NIPT screen. These data demonstrate that trisomy risk scores were accurately obtained from array-based analysis.

[000200] In addition, array data decreases the assay variability by approximately two-fold. Accuracy, as measured by decreasing assay variability, was enhanced for array-based hybridization analysis compared to sequencing-based analysis. The median assay variability for array sequence detection showed a nearly two-fold improvement over next-generation-sequencing-based detection (0.051 versus 0.099; p-value < 0.0001) (See FIG. 14). The bars of the histogram show the number of ligation assays (y-axis) that share a specific range of assay variability (x-axis). Array data is plotted in dark gray, sequencing data is plotted in white. Where the two populations of data overlap, the bars are light gray. The array-based quantified ligation products have significantly lower assay variability, where lower assay variability is better.

[000201] Arrays can be utilized to lower fetal fraction variability: The fraction of fetal DNA in plasma impacts the accuracy of the testing. The methods of the present invention measures, reports, and leverages fetal fraction using the FORTE algorithm to provide highly sensitive

results with a low false positive rate (Sparks, et al., Am J Obstet Gynecol., 206(4):319.e1-9 (2012)). FORTE calculated fetal fractions are extremely reproducible between array-based and sequencing-based analysis data (R^2 > 0.99). Because the array is able to simultaneously measure a larger number of ligation assays than multiplexed sequencing, fetal fraction estimates using array-based analysis are more precise with a median relative standard error of 0.013 compared to 0.021.

[000202] The data presented in the Examples show that two key sources of data variability are lower for arrays compared to next generation sequencing: 1) variation of the measured fetal fraction using polymorphic assays (fetal fraction variability) and 2) the variation of non-polymorphic assays across samples (assay variability). By including more polymorphic assays, an array-based ligation product prenatal test has been engineered that provides lower fetal-fraction variability resulting in better precision. By lowering assay variability, smaller aneuploidy changes can be measured, such as the smaller changes observed in low fetal fraction samples. The data demonstrates that the array platform is capable of reliable aneuploidy assessment. Moreover, array imaging is a rapid process and the turnaround time for sample quantitation is reduced to less than a minute per sample. The specific detection system employed in these Examples (GENETITAN™ Multi-Channel Instrument) can image >90 arrays per machine-hour. In contrast, even when samples are multiplexed in groups of 96 samples per lane, the HISEQ™ 2500 sequencing system has a throughput of 15 samples per machine-hour. This decrease in machine time and complexity translates directly to reductions in capital costs when array sequence detection is used.

[000203] The sequence based analysis leverages sample multiplexing in order to achieve economically efficient use of available sequence capacity. However, without normalization, a single sample can consume the majority of sequence reads in a flow-cell, reducing the reads available for determining trisomy risk in the remaining samples. In order to multiplex samples accurately, laborious and expensive processes to normalize quantities of input DNA are

required. Even when efforts are made to equalize sample input, as was reported in a recent study, a four-fold variation in the median reads per-sample was observed for a 12-plex reaction (Jensen, et al., PLoS ONe, 8(3):e57381 (Jul 2014). Array-based NIPT approaches require no sample multiplexing. Instead, each sample is hybridized individually to a single array. Processing throughput is enhanced by physically connecting, *e.g.,* 384 arrays onto a single multi-array plate for convenient high throughput handling. Because each sample is handled individually, time is saved and cost is reduced.

[000204] While this invention is satisfied by aspects in many different forms, as described in detail in connection with preferred aspects of the invention, it is understood that the present disclosure is to be considered as exemplary of the principles of the invention and is not intended to limit the invention to the specific aspects illustrated and described herein. Numerous variations may be made by persons skilled in the art without departure from the spirit of the invention. The scope of the invention will be measured by the appended claims and their equivalents. The abstract and the title are not to be construed as limiting the scope of the present invention, as their purpose is to enable the appropriate authorities, as well as the general public, to quickly determine the general nature of the invention. In the claims that follow, unless the term “means” is used, none of the features or elements recited therein should be construed as means-plus-function limitations pursuant to 35 U.S.C. §112, ¶6.

What is claimed is:

1. An assay method for providing a statistical likelihood of a fetal copy number variation comprising:

providing a maternal sample comprising maternal and fetal cell free DNA;

interrogating at least 48 non-polymorphic loci from a first target genomic region using at least two fixed sequence oligonucleotides comprising a region complementary to a locus in the first target genomic region wherein one of the fixed sequence oligonucleotides comprises a first capture region, a first label binding region, and two restriction sites;

interrogating at least 48 non-polymorphic loci from a second target genomic region using at least two fixed sequence oligonucleotides comprising a region complementary to a locus in the second target genomic region wherein one of the fixed sequence oligonucleotides comprises the first capture region, a second label binding region, and two restriction sites;

ligating the fixed sequence oligonucleotides;

amplifying the ligated fixed sequence oligonucleotides to create amplicons;

cleaving the amplicons to create cleaved amplicons;

detecting the cleaved amplicons from the first and second target genomic regions via hybridization of the first capture regions of the cleaved amplicons to an array comprising capture probes complementary to the first capture regions, wherein the cleaved amplicons from the first and second genomic target regions hybridize competitively to the capture probes complementary to the first capture regions;

quantifying the capture regions of the cleaved amplicons to determine a relative frequency of the interrogated non-polymorphic loci from the first and second genomic target regions by detecting the first and second label binding regions;

estimating the relative frequency of the first and second target genomic regions based on the relative frequency of the first and second label binding regions;

interrogating at least 48 polymorphic loci from at least one target genomic region different from the first and second target genomic regions using at least three fixed sequence allele-specific oligonucleotides for each polymorphic locus, wherein two of the at least three allele-specific oligonucleotides comprises a sequence complementary to one allele at a polymorphic locus, a capture region specific for each polymorphic locus, a different label binding region for each allele at the polymorphic locus, and two restriction sites;

ligating the fixed sequence oligonucleotides specific to the polymorphic loci;

amplifying the ligated fixed sequence oligonucleotides specific to the polymorphic loci to create amplicons;

cleaving the amplicons specific to the polymorphic loci to create cleaved amplicons;

detecting cleaved amplicons from the polymorphic loci via competitive hybridization of the capture regions of the cleaved amplicons to capture regions on the array;

quantifying the alleles of the polymorphic loci by detecting the different label binding regions for each allele on the cleaved amplicons to determine the fraction of fetal DNA in the sample;

determining the fraction of fetal DNA from the quantified alleles; and

calculating a statistical likelihood of a fetal copy number variation in the maternal sample using the estimated relative frequency of the first and second genomic regions in the sample and the fraction of fetal DNA.

2. The assay method of claim 1, wherein three or more synthetic fixed sequence oligonucleotides are used to interrogate each non-polymorphic locus in the first and second target genomic regions.

3. The assay method of claim 1, wherein the amplification steps comprise universal amplification using the polymerase chain reaction.

4. The assay method of claim 1, further comprising:

identifying low frequency alleles from the quantified alleles of the polymorphic loci where the maternal DNA is homozygous and the fetal DNA is heterozygous; and

using the low frequency alleles from the polymorphic loci to determine the fraction of fetal DNA.

5. The assay method of claim 1, wherein the first genomic region is a chromosome.

6. The assay method of claim 1, wherein the second genomic region is a chromosome.

7. The assay method claim 1, wherein the first genomic region is a sub-chromosomal region.

8. The assay method claim 1, wherein the second genomic region is a sub-chromosomal region.

9. The assay method of claim 1, wherein the label binding regions of the fixed oligonucleotides comprise a label.

10. The assay method of claim 1, wherein the label binding regions of the fixed oligonucleotides comprise a sequence complementary to an oligonucleotide comprising a label.

11. The assay method of claim 1, wherein the polymorphic loci are autosomal loci.

12. The assay method of claim 1, wherein the interrogating steps further comprise use of bridging oligonucleotides.

13. An assay method for determining a likelihood of a fetal aneuploidy comprising the steps of:

providing a maternal sample comprising maternal and fetal cell free DNA;

introducing at least fifty first sets of two or more fixed sequence oligonucleotides complementary to a set of non-polymorphic loci in a first target genomic region in the maternal sample under conditions that allow a complementary region of each fixed sequence oligonucleotide to specifically hybridize to the non-polymorphic loci, wherein at least one of the fixed sequence oligonucleotides of each set comprises a universal primer site, a first capture region, a first label binding region, and two restriction sites;

introducing at least fifty second sets of two or more fixed sequence oligonucleotides complementary to a set of non-polymorphic loci in a second target genomic region in the maternal sample under conditions that allow a complementary region of each fixed sequence oligonucleotide to specifically hybridize to the non-polymorphic loci, wherein at least one of the fixed sequence oligonucleotides of each set comprises a universal primer site, a first capture region, a second label binding region, and two restriction sites;

introducing two or more third sets of three or more fixed sequence oligonucleotides complementary to a set of polymorphic loci in the maternal sample under conditions that allow a complementary region of each fixed sequence oligonucleotide to specifically hybridize to a polymorphic locus, wherein at least two of the three fixed sequence oligonucleotides of each set comprises a universal primer site, a sequence complementary to one allele at the polymorphic locus, a different label binding region for each allele at the polymorphic locus, two restriction

sites, and a capture region, wherein the capture region for each polymorphic locus is different from the capture region for every other polymorphic locus and different from the first capture region;

hybridizing the first, second and third sets of fixed sequence oligonucleotides to the first and second target genomic regions and polymorphic loci;

extending at least one of the hybridized fixed sequence oligonucleotides of the first and second sets to form adjacently hybridized fixed sequence oligonucleotides;

ligating the adjacently hybridized first and second fixed sequence oligonucleotides from the first and second sets to create ligation products;

amplifying the ligation products using the universal primer sites to create amplicons;

cleaving the amplicons to create cleaved amplicons;

applying the cleaved amplicons to an array, wherein the array comprises first capture probes complementary to the first capture regions on the cleaved amplicons from the first and second target genomic regions, and wherein the array comprises capture probes complementary to the capture regions on the cleaved amplicons from each polymorphic locus;

hybridizing the first capture regions of the cleaved amplicons from the first and second target genomic regions to first capture probes on an array;

hybridizing the capture regions of the cleaved amplicons from the polymorphic loci to capture probes on the array;

detecting the hybridized cleaved amplicons;

quantifying a relative frequency of each allele from the polymorphic loci by detecting the different label binding regions for each allele on the cleaved amplicons to determine a percent fetal cell free DNA;

determining the percent of fetal cell free;

quantifying a relative frequency of cleaved amplicons corresponding to loci from the first target genomic region and a relative frequency of cleaved amplicons corresponding to loci from the second target genomic region by detecting the first and second label binding regions; and

computing a likelihood of a fetal aneuploidy using the relative frequency of cleaved amplicons corresponding to loci from the first and second target genomic regions and the percent fetal cell free DNA.

14. The assay method of claim 13, further comprising:

identifying low frequency alleles from the quantified alleles of the polymorphic loci where the maternal DNA is homozygous and the fetal DNA is heterozygous; and

using the low frequency alleles from the polymorphic loci to determine the fraction of fetal DNA.

15. The assay method of claim 13, further comprising comparing the relative frequencies of the cleaved amplicons corresponding to loci from the first and second target genomic regions and adjusting the likelihood of a fetal aneuploidy based on the percent fetal cell-free DNA.

16. The assay method of claim 13 wherein the relative frequencies of the cleaved amplicons corresponding to the loci from the first target genomic region are summed and the relative frequencies of the cleaved amplicons corresponding to the loci from the second target genomic region are summed, and the sum from the cleaved amplicons corresponding to the loci from the first target genomic region are compared to the sum from the cleaved amplicons corresponding to the loci from the second target genomic region to calculate a target genomic region ratio.

17. The assay method of claim 13, wherein the polymorphic loci are autosomal loci.

18. The assay method of claim 13, wherein the cleaved amplicons from loci of the first and second target genomic regions and the cleaved amplicons from the polymorphic loci are amplified in a single vessel.

19. The assay method of claim 13, wherein the label binding regions of the fixed oligonucleotides comprise a label.

20. The assay method of claim 13, wherein the label binding regions of the fixed oligonucleotides comprise a sequence complementary to an oligonucleotides comprising a label.