Abstract

The present invention provides apparatus and methods for enriching components or cells from a sample and conducting genetic analysis, such as SNP genotyping to provide diagnostic results for fetal disorders or conditions.
Obtain a mixed sample

Enrich fetal cells or fetal DNA

Obtain a reference sample

Extract genomic DNA

Perform DNA Pre-amplification

Detect SNP(s)

Estimate fetal/maternal ratio

Determine Aneuploidy

Figure 1
Figure 2A
Figure 2B
Microposts and cells

Figure 3A

Antibody coated posts

Figure 3B

Figure 3C
Figure 4

- Sample in
- Sample out
- Magnetic flux field
- Polymer matrix with hard/soft magnetic material
- Supra or para magnetic particle
- Monoclonal antibody to transferrin receptor
- Target cell
- Transferrin receptor
- Silicon
Figure 5
Typical allele strengths at different loci.
First row – maternal sample.
Second row – mixed sample containing maternal and fetal DNA.
Third row – mixed sample assuming fetal trisomy at Loci 1 and 2.
Excess signal from the extra chromosome copy is indicated in dashed red.

Figure 6
Concept of rank ordering of allele strengths. Assuming a bi-allelic maternal sample, a third allele significantly above the noise level can be identified as paternal.

Figure 7
Histogram of paternal allele strength, normalized to average of maternal alleles.

Figure 8
Figure 10A-F (Blue = nucleus, Red = X chromosome, Green = Y chromosome)
FIGURE 13
FIGURE 15
ASSAY

Whole Genome

Single Gene Disorder

Trisomy 13, 18, 21

Downs

Few target cells, high maternal bkg

Several target cells, high maternal bkg

One pure cell

Several pure cells

FIGURE 16
Fetal Cell ID

Fetal Cell isolation

Genome analysis

Anomaly Detection

Affymetrix array cord vs maternal blood—ID off in maternal/on in fetal

Biomarker test

Hemoglobin markers coincident with DYZ

Select sites with fetal cell presence

Whole genome amplification

CGH array

Ultra deep sequencing

Mass spectrometry

Solexa 454

Allele-specific SNP panel & Taqman

Purification Methods

FIGURE 18
USE OF HIGHLY PARALLEL SNP GENOTYPING FOR FETAL DIAGNOSIS

BACKGROUND OF THE INVENTION

This application claims the benefit of U.S. Provisional Application No. 60/804,816, filed Jun. 14, 2006, which application is incorporated herein by reference.

Analysis of specific cells can give insight into a variety of diseases. These analyses can provide non-invasive tests for detection, diagnosis and prognosis of diseases, thereby eliminating the risk of invasive diagnosis. For instance, social developments have resulted in an increased number of prenatal tests. However, the available methods today, amniocentesis and chorionic villus sampling (CVS) are potentially harmful to the mother and to the fetus. The rate of miscarriage for pregnant women undergoing amniocentesis is increased by 0.5-1%, and that figure is slightly higher for CVS. Because of the inherent risks posed by amniocentesis and CVS, these procedures are offered primarily to older women, i.e., those over 35 years of age, who have a statistically greater probability of bearing children with congenital defects. As a result, a pregnant woman at the age of 35 has to balance an average risk of 0.5-1% to induce an abortion by amniocentesis against an age related probability for trisomy 21 of less than 0.3%.

Some non-invasive methods have already been developed to diagnose specific congenital defects. For example, maternal serum alpha-fetoprotein, and levels of unconjugated estriol and human chorionic gonadotropin can be used to identify a proportion of fetuses with Down’s syndrome, however, these tests are not one hundred percent accurate. Similarly, ultrasonography is used to determine congenital defects involving neural tube defects and limb abnormalities, but is useful only after fifteen weeks’ gestation.

The methods of the present invention allow for the detection of fetal cells and fetal abnormalities when fetal cells are mixed with a population of maternal cells, even when the maternal cells dominate the mixture.

SUMMARY OF THE INVENTION

The presence of fetal cells within the blood of pregnant women offers the opportunity to develop a prenatal diagnostic that replaces amniocentesis and thereby eliminates the risk of today’s invasive diagnosis. However, fetal cells represent a small number of cells against the background of a large number of maternal cells in the blood which make the analysis time consuming and prone to error. Current technologies and protocols for highly parallel SNP detection with DNA microarray readout result in inaccurate calls when there are too few starting DNA copies or when a particular allele represents a small fraction in the population of input DNA molecules.

The present invention relates to methods for detecting a fetal abnormality by determining the ratio of the abundance of one or more maternal alleles to the abundance of one or more paternal alleles in the genomic DNA of a sample. The genomic region includes a single nucleotide polymorphism (SNP), which can preferably be an informative SNP. The SNP can be detected by methods that include using a DNA microarray, bead microarray, or high throughput sequencing. In some embodiments, determining the ratio involves detecting an abundance of a nucleotide base at a SNP position. In other embodiments, determining the ratio also comprises calculating error rate based amplification. Prior to determining the abundance of allele(s), the sample can be enriched for fetal cells.

The method of detection is provided by highly parallel SNP detection that can be used to determine the ratios of abundance of maternal and paternal alleles at a plurality of genomic regions present in the sample. In some embodiments, the ratios of abundance are determined in at least 100 genomic regions, which can comprise a single locus, different loci, a single chromosome, or different chromosomes. In some embodiments, a first genomic region (SNP) analyzed is in a genomic region suspected of being trisomic or is trisomic and a second genomic region (SNP) analyzed is in a non-trisomic region or a region suspected of being non-trisomic. The ratio of alleles in the first genomic region can then be compared to the ratio of alleles in the second genomic region, and in some embodiments, the comparison is made by determining the difference in the means of the ratios in the first and second genomic regions. An increase in paternal abundance can be indicative of paternal trisomy, while an increase in maternal abundance can be indicative of maternal trisomy. Alternatively, an increase in paternal abundance or maternal abundance of one or more alleles is indicative of partial trisomy. The first and second genomic regions can be on the same or different chromosomes.

In an embodiment, the invention provides for a method for detecting a fetal abnormality comprising comparing an abundance of one or more maternal alleles in a first genomic region in a maternal blood sample, where said genomic region is suspected of trisomy with an abundance of one or more maternal alleles in a second genomic region in said blood sample wherein said second genomic region is non-trisomic. Up to 20 ml of blood can be used to detect the fetal abnormality. The first genomic region that is suspected of trisomy and the second genomic region that is a non-trisomic region can each be present on chromosomes 13, 18, 21 and on the X chromosome.

In some embodiments, a ratio of the abundance of the maternal alleles in the first genomic region to the abundance of the maternal alleles in the second genomic region can be determined and compared to a second ratio obtained for a control sample. The control sample can comprise a diluted portion of the maternal sample, which can be diluted by a factor of at least 1,000.

In some embodiments, detecting the fetal abnormality further involves estimating the number of fetal cells present in the maternal sample. This can be performed by, e.g., ranking the alleles detected according to their abundance. The ranking can then be used to determine an abundance of one or more paternal alleles. In some embodiments, data models can be fitted for optimal detection of aneuploidy. The methods herein can be used to identify monoploidy, triploidy, tetraploidy, pentaploidy and other multiples of the normal euploid state. For example, the data models can be used to determine estimates for the fraction of fetal cells present in a sample and for detecting a fetal abnormality or condition.
In some embodiments, the abundance of one or more paternal alleles can be compared to the abundance of the maternal alleles at one or more genetic regions. In other embodiments, one or more ratios of the abundance of the paternal allele(s) to the abundance of the maternal allele(s) at one or more genetic regions can be compared with an estimate fraction of fetal cells. A statistical analysis can be performed on the one or more ratios of the abundance of paternal alleles to the abundance of the maternal alleles to determine the presence of fetal DNA in the sample with a level of confidence that exceeds 90%.

**INCORPORATION BY REFERENCE**

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

**BRIEF DESCRIPTION OF THE DRAWINGS**

The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

**FIG. 1** illustrates an overview of the process of the invention.

**FIGS. 2A-2D** illustrates one embodiment of a size-based separation module.

**FIGS. 3A-3C** illustrates one embodiment of an affinity separation module.

**FIG. 4** illustrates one embodiment of a magnetic separation module.

**FIG. 5** illustrates an overview for a typical parallel SNP genotyping assay.

**FIG. 6** illustrates the types of SNP calls that result depicting allele strengths at different loci.

**FIG. 7** illustrates the concept of rank ordering of allele strengths.

**FIG. 8** illustrates a histogram of paternal allele strength normalized relative to maternal alleles.

**FIGS. 9A-9B** illustrate cell smears of the product and waste fractions.

**FIGS. 10A-10F** illustrate isolated fetal cells confirmed by the reliable presence of male Y chromosome.

**FIG. 11** illustrates trisomy 21 pathology in an isolated fetal nucleated red blood cell.

**FIGS. 12A-12D** illustrate various embodiments of a size-based separation module.

**FIG. 13** illustrates the detection of single copies of a fetal cell genome by qPCR.

**FIG. 14** illustrates detection of single fetal cells in binned samples by SNP analysis.

**FIG. 15** illustrates a method of trisomy testing. The trisomy 21 screen is based on scoring of target cells obtained from maternal blood. Blood is processed using a cell separation module for hemoglobin enrichment (CSM-HE). Isolated cells are transferred to slides that are first stained and subsequently probed by FISH. Images are acquired, such as from bright field or fluorescent microscopy, and scored. The proportion of trisomic cells of certain classes serves as a classifier for risk of fetal trisomy 21. Fetal genome identification can be performed using assays such as: (1) STR markers; (2) qPCR using primers and probes directed to loci, such as the multi-repeat DYZ locus on the Y-chromosome; (3) SNP detection; and (4) CGH (comparative genome hybridization) array detection.

**FIG. 16** illustrates assays that can produce information on the presence of aneuploidy and other genetic disorders in target cells. Information on aneuploidy and other genetic disorders in target cells may be acquired using technologies such as: (1) a CGH array established for chromosome counting, which can be used for aneuploidy determination and/or detection of intra-chromosomal deletions; (2) SNP/taqman assays, which can be used for detection of single nucleotide polymorphisms; and (3) ultra-deep sequencing, which can be used to produce partial or complete genome sequences for analysis.

**FIG. 17** illustrates methods of fetal diagnostic assays. Fetal cells are isolated by CSM-HE enrichment of target cells from blood. The designation of the fetal cells may be confirmed using techniques comprising FISH staining (using slides or membranes and optionally an automated detector), FACS, and/or binning. Binning may comprise distribution of enriched cells across wells in a plate (such as a 96 or 384 well plate), microencapsulation of cells in droplets that are separated in an emulsion, or by introduction of cells into microarrays of nanofluidic bins. Fetal cells are then identified using methods that may comprise the use of biomarkers (such as fetal (gamma) hemoglobin), allele-specific SNP panels that could detect fetal genome DNA, detection of differentially expressed maternal and fetal transcripts (such as Affymetrix chips), or primers and probes directed to fetal specific loci (such as the multi-repeat DYZ locus on the Y-chromosome). Binning sites that contain fetal cells are then analyzed for aneuploidy and/or other genetic defects using a technique such as CGH array detection, ultra deep sequencing (such as Solexa, 454, or mass spectrometry), STR analysis, or SNP detection.

**FIG. 18** illustrates methods of fetal diagnostic assays, further comprising the step of whole genome amplification prior to analysis of aneuploidy and/or other genetic defects.

**DETAILED DESCRIPTION OF THE INVENTION**

The methods herein are used for detecting the presence and condition of fetal cells in a mixed sample wherein the fetal cells are at a concentration of less than 90, 80, 70, 60, 50, 40, 30, 20, 10, 5 or 1% of all cells in the sample at a concentration less than 1, 2, 4, 1, 10, 1, 100, 1, 1000, 1, 10,000, 1, 100,000, 1, 000,000, 1, 10,000, 000 or 1,100,000,000 of all cells in the sample.

**FIG. 1** illustrates an overview of the methods and systems herein.
In step 100, a sample to be analyzed for rare cells (e.g. fetal cells) is obtained from an animal. Such animal can be suspected of being pregnant, pregnant, or one that has been pregnant. Such sample can be analyzed by the systems and methods herein to determine a condition in the animal or fetus of the animal. In some embodiments, the methods herein are used to detect the presence of a fetus, sex of a fetus, or condition of the fetus. The animal from whom the sample is obtained can be, for example, a human or a domesticated animal such as a cow, chicken, pig, horse, rabbit, dog, cat, or goat. Samples derived from an animal or human include, e.g., whole blood, sweat, tears, ear flow, sputum, lymph, bone marrow suspension, lymph, urine, saliva, semen, vaginal flow, cerebrospinal fluid, brain fluid, ascites, milk, secretions of the respiratory, intestinal or genitourinary tracts fluid.

To obtain a blood sample, any technique known in the art may be used, e.g., a syringe or other vacuum suction device. A blood sample can be optionally pre-treated or processed prior to enrichment. Examples of pre-treatment steps include the addition of a reagent such as a stabilizer, a preservative, a fixant, a lysing reagent, a diluent, an anti-apoptotic reagent, an anti-coagulation reagent, an anti-thrombotic reagent, magnetic property regulating reagent, a buffering reagent, an osmolality regulating reagent, a pH regulating reagent, and/or a cross-linking reagent.

When a blood sample is obtained, a preservative such an anti-coagulation agent and/or a stabilizer is often added to the sample prior to enrichment. This allows for extended time for analysis/detection. Thus, a sample, such as a blood sample, can be enriched and/or analyzed under any of the methods and systems herein within 1 week, 6 days, 5 days, 4 days, 3 days, 2 days, 1 day, 12 hrs, 6 hrs, 3 hrs, 2 hrs, or 1 hr from the time the sample is obtained.

A blood sample can be combined with an agent that selectively lyzes one or more cells or components in a blood sample. For example, fetal cells can be selectively lysed releasing their nuclei when a blood sample including fetal cells is combined with deionized water. Such selective lysis allows for the subsequent enrichment of fetal nuclei using, e.g., size or affinity based separation. In another example, platelets and/or nucleated red blood cells are selectively lysed to generate a sample enriched in nucleated cells, such as fetal nucleated red blood cells (fRBC) and material red nucleated blood cells (mRBC). The fRBCs can subsequently be separated from the mRBCs using, e.g., antigen-1 affinity or differences in hemoglobin.

When obtaining a sample from an animal (e.g., blood sample), the amount can vary depending upon animal size, its gestation period, and the condition being screened. Up to 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 mL of a sample is obtained. The volume of sample obtained can be 1-50, 2-40, 3-30, or 4-20 mL. Alternatively, more than 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 mL of a sample is obtained.

To detect fetal abnormality, a blood sample can be obtained from a pregnant animal or human within 36, 24, 22, 20, 18, 16, 14, 12, 10, 8, 6 or 4 weeks of gestation.

In step 101, a reference or control sample is obtained by any means known in the art. A reference sample is any sample that consists essentially of, or only of, non-fetal cells or non-fetal DNA. A reference sample is preferably a maternal only cell or DNA sample. In some embodiments, a reference sample is a maternal only blood sample. When obtaining a reference sample such as a maternal blood sample from a pregnant female, or one suspected of being pregnant or the sample can be diluted enough to ensure that <1 fetal cell is expected in the sample. Dilution can be by a factor of about 10 to 1000 fold, or by a factor of greater than 5, 10, 50, 100, 200, 500 to 1000 fold. Alternatively, offsite blood cells can be obtained from the same organism from whom the mixed sample is obtained. In some cases, the reference sample is obtained by deleting a portion of the mixed sample.

In step 102, when the sample to be tested or analyzed is a mixed sample (e.g. maternal blood sample), it is enriched for rare cells or rare DNA (e.g. fetal cells, fetal DNA or fetal nuclei) using one or more methods known in the art or disclosed herein. Such enrichment increases the ratio of fetal cells to non-fetal cells, the concentration of fetal DNA to non-fetal DNA, and/or the concentration of fetal cells in volume per total volume of the mixed sample.

In some embodiments, enrichment occurs by selective lysis as described above. For example, enucleated cells may be selectively lysed prior to subsequent enrichment steps or fetal nucleated cells may be selectively lysed prior to separation of the fetal nuclei from other cells and components in the sample.

In some embodiments, enrichment of fetal cells or fetal nuclei occurs using one or more size-based separation modules. Size-based separation modules include filtration modules, sieves, matrixes, etc., including those disclosed in International Publication Nos. WO 2004/113877, WO 2004/0144651, and US Application Publication No. 2004/011156.

In some embodiments, a size-based separation module includes one or more arrays of obstacles that form a network of gaps. The obstacles are configured to direct particles (e.g. cells or nuclei) as they flow through the array/network of gaps into different directions or outlets based on the particle's hydrodynamic size. For example, as a blood sample flows through an array of obstacles, nucleated cells or cells having a hydrodynamic size larger than a predetermined size, e.g., 8 microns, are directed to a first outlet located on the opposite side of the array of obstacles from the fluid flow inlet, while the enucleated cells or cells having a hydrodynamic size smaller than a predetermined size, e.g., 8 microns, are directed to a second outlet also located on the opposite side of the array of obstacles from the fluid flow inlet.

An array can be configured to separate cells smaller than a predetermined size from those larger than a predetermined size by adjusting the size of the gaps, obstacles, and offset in the period between each successive row of obstacles. For example, in some embodiments, obstacles and/or gaps between obstacles can be up to 10, 20, 50, 70, 100, 120, 150, 170, or 200 microns in length or about 2, 4, 6, 8 or 10 microns in length. In some embodiments, an array for size-based separation includes more than 100, 500, 1,000, 5,000, 10,000, 50,000 or 100,000 obstacles that are arranged into more than 10, 20, 50, 100, 200, 500, or 1000 rows. Preferably, obstacles in a first row of obstacles are offset from a previous (upstream) row of obstacles by up to 50% the period of the previous row of obstacles. In some
embodiments, obstacles in a first row of obstacles are offset from a previous row of obstacles by up to 45, 40, 35, 30, 25, 20, 15 or 10% the period of the previous row of obstacles. Furthermore, the distance between a first row of obstacles and a second row of obstacles can be up to 10, 20, 50, 70, 100, 120, 150, 170 or 200 microns. A particular offset can be continuous (repeating for multiple rows) or non-continuous.

In some embodiments, a separation module includes multiple discrete arrays of obstacles fluidly coupled such that they are in series with one another. Each array of obstacles has a continuous offset. But each subsequent (downstream) array of obstacles has an offset that is different from the previous (upstream) offset. Preferably, each subsequent array of obstacles has a smaller offset that the previous array of obstacles. This allows for a refinement in the separation process as cells migrate through the array of obstacles. Thus, a plurality of arrays can be fluidly coupled in series or in parallel, (e.g., more than 2, 4, 6, 8, 10, 20, 30, 40, 50). Fluidly coupling separation modules (e.g., arrays) in parallel allows for high-throughput analysis of the sample, such that at least 1, 2, 5, 10, 20, 50, 100, 200, or 500 mL per hour flows through the enrichment modules or at least 1, 5, 10, or 50 million cells per hour are sorted or flow through the device.

FIGS. 2A-2D illustrates an example of a size-based separation module. Obstacles (which may be of any shape) are coupled to a flat substrate to form an array of gaps. A transparent cover or lid may be used to cover the array. The obstacles form a two-dimensional array with each successive row shifted horizontally with respect to the previous row of obstacles, where the array of obstacles directs component having a hydrodynamic size smaller than a predetermined size in a first direction and component having a hydrodynamic size larger that a predetermined size in a second direction. The flow of sample into the array of obstacles can be aligned at a small angle (flow angle) with respect to a line-of-sight of the array. Optionally, the array is coupled to an infusion pump to perfuse the sample through the obstacles. The flow conditions of the size-based separation module described herein are such that cells are sorted by the array with minimal damage. This allows for downstream analysis of intact cells and intact nuclei to be more efficient and reliable.

In one embodiment, a size-based separation module comprises an array of obstacles configured to direct fetal cells larger than a predetermined size to migrate along a line-of-sight within the array towards a first outlet or bypass channel leading to a first outlet, while directing cells and analytes smaller than a predetermined size through the array of obstacles in a different direction towards a second outlet.

A variety of enrichment protocols may be utilized although, in most embodiments, gentle handling of the cells is needed to reduce any mechanical damage to the cells or their DNA. This gentle handling also preserves the small number of fetal cells in the sample. Integrity of the nucleic acid being evaluated is an important feature to feature the distinction between the genomic material from the fetal cells and other cells in the sample. In particular, the enrichment and separation of the fetal cells using the arrays of obstacles produces gentle treatment which minimizes cellular damage and maximizes nucleic acid integrity permitting exceptional levels of separation and the ability to subsequently utilize various formats to very accurately analyze the genome of the cells which are present in the sample in extremely low numbers.

[0049] In some embodiments, enrichment of fetal cells occurs using one or more capture modules that selectively inhibit the mobility of one or more cells of interest. Preferably a capture module is fluidly coupled downstream to a size-based separation module. Capture modules can include a substrate having multiple obstacles that restrict the movement of cells or analytes greater than a predetermined size. Examples of capture modules that inhibit the migration of cells based on size are disclosed in U.S. Pat. Nos. 5,837,115 and 6,692,952.

[0050] In some embodiments, a capture module includes a two dimensional array of obstacles that selectively filters or captures cells or analytes having a hydrodynamic size greater than a particular gap size, e.g., predetermined size. Arrays of obstacles adapted for separation by capture can include obstacles having one or more shapes and can be arranged in a uniform or non-uniform order. In some embodiments, a two-dimensional array of obstacles is staggered such that each subsequent row of obstacles is offset from the previous row of obstacles to increase the number of interactions between the analytes being sorted (separated) and the obstacles.

[0051] Another example of a capture module is an affinity-based separation module. An affinity-based separation module captures analytes or cells of interest based on their affinity to a structure or particle as opposed to their size. One example of an affinity-based separation module is an array of obstacles that are adapted for complete sample flow through, but for the fact that the obstacles are covered with binding moieties that selectively bind one or more analytes (e.g., cell population) of interest (e.g., red blood cells, fetal cells, or nucleated cells) or analytes not-of-interest (e.g., white blood cells). Binding moieties can include e.g., proteins (e.g., ligands/receptors), nucleic acids having complementary counterparts in retained analytes, antibodies, etc. In some embodiments, an affinity-based separation module comprises a two-dimensional array of obstacles covered with one or more antibodies selected from the group consisting of: anti-CD71, anti-CD235a, anti-CD36, anti-carbohydrates, anti-selectin, anti-CD45, anti-GPA, and anti-antigen-I.

[0052] FIG. 3A illustrates a path of a first analyte through an array of posts wherein an analyte that does not specifically bind to a post continues to migrate through the array, while an analyte that does bind a post is captured by the array. FIG. 3B is a picture of antibody coated posts. FIG. 3C illustrates coupling of antibodies to a substrate (e.g., obstacles, side walls, etc.) as contemplated by the present invention. Examples of such affinity-based separation modules are described in International Publication No. WO 2004/029221.

[0053] In some embodiments, a capture module utilizes a magnetic field to separate and/or enrich one or more analytes (cells) that has a magnetic property or magnetic potential. For example, red blood cells which are slightly diamagnetic (repelled by magnetic field) in physiological conditions can be made paramagnetic (attracted by magnetic field) by deoxygenation of the hemoglobin into methemoglobin. This magnetic property can be achieved through physical or
chemical treatment of the red blood cells. Thus, a sample containing one or more red blood cells and one or more non-red blood cells can be enriched for the red blood cells by first inducing a magnetic property and then separating the above red blood cells from other analytes using a magnetic field (uniform or non-uniform). For example, a maternal blood sample can flow first through a size-based separation module to remove enucleated cells and cellular components (e.g., analytes having a hydrodynamic size less than 6 μm) based on size. Subsequently, the enriched nucleated cells (e.g., analytes having a hydrodynamic size greater than 6 μm) white blood cells and nucleated red blood cells are treated with a reagent, such as CO₂, N₂, or NaNO₃, that changes the magnetic property of the red blood cells’ hemoglobin. The treated sample then flows through a magnetic field (e.g., a column coupled to an external magnet), such that the paramagnetic analytes (e.g., red blood cells) will be captured by the magnetic field while the white blood cells and any other non-red blood cells will flow through the device to result in a sample enriched in nucleated red blood cells (including fRBC’s). Additional examples of magnetic separation modules are described in U.S. application Ser. No. 11/323,971, filed Dec. 29, 2005 entitled “Devices and Methods for Magnetic Enrichment of Cells and Other Particles” and U.S. application Ser. No. 11/227,904, filed Sep. 15, 2005, entitled “Devices and Methods for Enrichment and Alteration of Cells and Other Particles”.

Subsequent enrichment steps can be used to separate the rare cells (e.g., fRBC’s) from the non-rare maternal nucleated red blood cells (non-RBC’s). In some embodiments, a sample enriched by size-based separation followed by affinity/magnetic separation is further enriched for rare cells using fluorescence activated cell sorting (FACS) or selective lysis of a subset of the cells (e.g. fetal cells). In some embodiments, fetal cells are selectively bound to an anti-antigen i binding moiety (e.g. an antibody) to separate them from the mmRBC’s. In some embodiments, fetal cells or fetal DNA is distinguished from non-fetal cells or non-fetal DNA by forcing the rare cells (fetal cells) to become apoptotic, thus condensing their nuclei and optionally ejecting their nuclei. Rare cells such as fetal cells can be forced into apoptosis using various means including subjecting the cells to hyperbaric pressure (e.g. 4% CO₂). The condensed nuclei can be detected and/or isolated for further analysis using any technique known in the art including DNA gel electrophoresis, in situ labeling of DNA nicked (terminal deoxynucleotidyl transferase (TdT)-mediated dUTP in situ nick labeling (also known as TUNEL) (Gavrieli, Y., et al. J. Cell Biol 119:493-501 (1992)) and ligation of DNA strand breaks having one or two-base 3’ overhangs (Taq polymerase-based in situ ligation). (Didenko V., et al. J. Cell Biol. 135:1369-76 (1996)).

In some embodiments, when the analyte desired to be separated (e.g., red blood cells or white blood cells) is not ferromagnetic or does not have a magnetic property, a magnetic particle (e.g., a bead) or compound (e.g., Fe³⁺) can be coupled to the analyte to give it a magnetic property. In some embodiments, a bead coupled to an antibody that selectively binds to an analyte of interest can be decorated with an antibody elected from the group of anti CD71 or CD75. In some embodiments a magnetic compound, such as Fe³⁺, can be coupled to an antibody such as those described above. The magnetic particles or magnetic antibodies herein may be coupled to any one or more of the devices described herein prior to contact with a sample or may be mixed with the sample prior to delivery of the sample to the device(s).

The magnetic field used to separate analytes/cells in any of the embodiments herein can uniform or non-uniform as well as external or internal to the device(s) herein. An external magnetic field is one whose source is outside a device herein (e.g., container, channel, obstacles). An internal magnetic field is one whose source is within a device contemplated herein. An example of an internal magnetic field is one where magnetic particles may be attached to obstacles present in the device (or manipulated to create obstacles) to increase surface area for analytes to interact with to increase the likelihood of binding. Analytes captured by a magnetic field can be released by demagnetizing the magnetic regions retaining the magnetic particles. For selective release of analytes from regions, the demagnetization can be limited to selected obstacles or regions. For example, the magnetic field can be designed to be electromagnetic, enabling turn-on and turn-off of the magnetic fields for each individual region or obstacle at will.

FIG. 4 illustrates an embodiment of a device configured for capture and isolation of cells expressing the transferrin receptor from a complex mixture. Monoclonal antibodies to CD71 receptor are readily available off-the-shelf and can be covalently coupled to magnetic materials, such as, but not limited to any conventional ferroparticles including ferrous doped polystyrene and ferroparticles or ferro-colloids (e.g., from Miltenyi or Dynal). The anti CD71 bound to magnetic particles is flowed into the device. The antibody coated particles are drawn to the obstacles (e.g., posts), floor, and walls and are retained by the strength of the magnetic field interaction between the particles and the magnetic field. The particles between the obstacles, and those loosely retained with the sphere of influence of the local magnetic fields away from the obstacles, are removed by a rinse.

One or more of the enrichment modules herein (e.g., size-based separation module(s) and capture module(s)) may be fluidly coupled in series or in parallel with one another. For example a first outlet from a separation module can be fluidly coupled to a capture module. In some embodiments, the separation module and capture module are integrated such that a plurality of obstacles acts both to deflect certain analytes according to size and direct them to a path different than the direction of analyte(s) of interest, and also as a capture module to capture, retain, or bind certain analytes based on size, affinity, magnetism or other physical property.

In any of the embodiments herein, the enrichment steps performed have a specificity and/or sensitivity ≥50, 60, 70, 80, 90, 95, 96, 97, 98, 99, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, 99.9 or 99.95% The retention rate of the enrichment module(s) herein is such that ≥50, 60, 70, 80, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 99.9 % of the analytes or cells of interest (e.g., nucleated cells or nucleated red blood cells or nucleated from red blood cells) are retained. Simultaneously, the enrichment modules are configured to remove ≥50, 60, 70, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 99.9 % of all unwanted analytes (e.g., red blood-platelet enriched cells) from a sample.

Any or all of the enrichment steps can occur with minimal dilution of the sample. For example, in some
embodiments the analytes of interest are retained in an enriched solution that is less than 50, 40, 30, 20, 10, 9.0, 8.0, 7.0, 6.0, 5.0, 4.5, 4.0, 3.5, 3.0, 2.5, 2.0, 1.5, 1.0, or 0.5 fold diluted from the original sample. In some embodiments, any or all of the enrichment steps increase the concentration of the analyte of interest (e.g., fetal cells), for example, by transferring them from the fluid sample to an enriched fluid sample (sometimes in a new fluid medium, such as a buffer). The new concentration of the analyte of interest may be at least 2, 4, 6, 8, 10, 20, 50, 100, 200, 500, 1,000, 2,000, 5,000, 10,000, 20,000, 50,000, 100,000, 200,000, 500,000, 1,000, 000, 2,000,000, 5,000,000, 10,000,000, 20,000,000, 50,000, 000, 100,000,000, 200,000,000, 500,000,000, 1,000,000, 000, 2,000,000,000, or 5,000,000,000 fold more concentrated than in the original sample. For example, a 10 times concentration increase of a first cell type out of a blood sample means that the ratio of first cell type/all cells in a sample is 10 times greater after the sample was applied to the apparatus herein. Such concentration can take a fluid sample (e.g., a blood sample) of greater than 10, 15, 20, 50, or 100 ml. total volume comprising rare components of interest, and it can concentrate such rare component of interest into a concentrated solution of less than 0.5, 1, 2, 3, 5, or 10 ml. total volume.

[0061] The final concentration of rare cells in relation to non-rare cells after enrichment can be about $10^{3}$, or $10^{4}$ to $10^{5}$. In some embodiments, the concentration of fetal cells to maternal cells may be up to $10^{4}$, i.e., 10,000,000 per ml, or as low as $10^{3}$, i.e., 1,000, or $10^{5}$, i.e., 100,000.

[0062] Thus, detection and analysis of the fetal cells can occur even if the non-fetal (e.g., maternal) cells are >50%, 60%, 70%, 80%, 90%, 95%, or 99% of all cells of a sample. In some embodiments, fetal cells are at a concentration of less than 1:2, 1:4, 1:10, 1:50, 1:100, 1:1000, 1:10,000, 1:100,000, 1:1,000,000, 1:10,000,000, or 1:100,000,000 of all cells in mixed sample to be analyzed or at a concentration of less than $10^{-3}$, $10^{-4}$, $10^{-5}$, $10^{-6}$, $10^{-7}$, or $10^{-8}$ cells/µl of the mixed sample. Over all, the number of fetal cells in a mixed sample (e.g., enriched sample) has up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 100 total fetal cells.

[0063] Enriched target cells (e.g., fnRBC) can be “binned” prior to analysis of the enriched cells (FIGS. 17 and 18). Binning is any process which results in the reduction of complexity and/or total number of the enriched cell output. Binning may be performed by any method known in the art or described herein. One method of binning the enriched cells is by serial dilution. Such dilution may be carried out using any appropriate platform (e.g., PCR wells, microtiter plates). Other methods include nanofluidic systems which separate samples into droplets (e.g., BioTrove, Raindance, Fluidigm). Such nanofluidic systems may result in the presence of a single cell present in a nanodroplet.

[0064] Binning may be preceded by positive selection for target cells including, but not limited to affinity binding (e.g., using anti-CD71 antibodies). Alternatively, negative selection of non-target cells may precede binning. For example, output from the size-based separation module may be passed through a magnetic hemoglobin enrichment module (MHEM) which selectively removes WBCs from the enriched sample.

[0065] For example, the possible cellular content of output from enriched maternal blood which has been passed through a size-based separation module (with or without further enrichment by processing the enriched sample through a MHEM) may consist of: 1) approximately 20 fnRBC; 2) 1,500 miniRBC; 3) 4,000-40,000 WBC; 4) 15×10⁶ RBC. If this sample is separated into 100 bins (PCR wells or other acceptable binning platform), each bin would be expected to contain: 1) 80 negative bins and 20 bins positive for one fnRBC; 2) 150 miniRBC; 3) 400-4,000 WBC; 4) 15×10⁶ RBC. If separated into 10,000 bins, each bin would be expected to contain: 1) 9,980 negative bins and 20 bins positive for one fnRBC; 2) 8,500 negative bins and 1,500 bins positive for one miniRBC; 3) 1-4 WBC; 4) 15×10⁶ RBC. One of skill in the art will recognize that the number of bins may be increased depending on experimental design and/or the platform used for binning. The reduced complexity of the binned cell populations may facilitate further genetic and cellular analysis of the target cells.

[0066] Analysis may be performed on individual bins to confirm the presence of target cells (e.g., fnRBC) in the individual bin. Such analysis may consist of any method known in the art, including, but not limited to, FISH, PCR, STR detection, SNP analysis, biomarker detection, and sequence analysis (FIGS. 17 and 18).

[0067] Fetal Biomarkers

[0068] In some embodiments fetal biomarkers may be used to detect and/or isolate fetal cells, after enrichment or after detection of fetal abnormality or lack thereof. For example, this may be performed by distinguishing between fetal and maternal nRBCs based on relative expression of a gene (e.g., DYS1, DYZ, CD-71, e- and ζ-globin) that is differentially expressed during fetal development. In preferred embodiments, biomarker genes are differentially expressed in the first and/or second trimester. “Differentially expressed,” as applied to nucleotide sequences or polypeptide sequences in a cell or cell nuclei, refers to differences in over/under-expression of that sequence when compared to the level of expression of the same sequence in another sample, a control or a reference sample. In some embodiments, expression differences can be temporal and/or cell-specific. For example, for cell-specific expression of biomarkers, differential expression of one or more biomarkers in the cell(s) of interest can be higher or lower relative to background cell populations. Detection of such difference in expression of the biomarker may indicate the presence of a rare cell (e.g., fnRBC) versus other cells in a mixed sample (e.g., background cell populations). In other embodiments, a ratio of two or more such biomarkers that are differentially expressed can be measured and used to detect rare cells.

[0069] In one embodiment, fetal biomarkers comprise differentially expressed hemoglobinins. Erythroblasts (nRBCs) are very abundant in the early fetal circulation, virtually absent in normal adult blood and by having a short finite lifespan, there is no risk of obtaining fnRBC which may persist from a previous pregnancy. Furthermore, unlike trophoblast cells, fetal erythroblasts are not prone to normal characteristics.

[0070] Yolk sac erythroblasts synthesize e-, ζ-, γ- and α-globins, these combine to form the embryonic hemoglobinins. Between six and eight weeks, the primary site of erythropoiesis shifts from the yolk sac to the liver, the three embryonic hemoglobinins are replaced by fetal hemoglobinin (HbF) as the predominant oxygen transport system, and e-
and ζ-globin production gives way to γ-, α- and β-globin production within definitive erythrocytes (Peschle et al., 1985). Hbf remains the principal hemoglobin until birth, when the second globin switch occurs and β-globin production accelerates.

Hemoglobin (Hb) is a heterodimer composed of two identical α-globin chains and two copies of a second globin. Due to differential gene expression during fetal development, the composition of the second chain changes from α-globin during early embryonic development (1 to 4 weeks of gestation) to γ-globin during fetal development (6 to 8 weeks of gestation) to β-globin in neonates and adults as illustrated in (Table 1).

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Relative expression of ε, γ and β in maternal and fetal RBCs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ε</td>
</tr>
<tr>
<td>1st trimester</td>
<td>Fetal</td>
</tr>
<tr>
<td></td>
<td>Maternal</td>
</tr>
<tr>
<td>2nd trimester</td>
<td>Fetal</td>
</tr>
<tr>
<td></td>
<td>Maternal</td>
</tr>
</tbody>
</table>

In the late-first trimester, the earliest time that fetal cells may be sampled by CVS, finRBCs contain, in addition to α-globin, primarily ε and γ-globin. In the early to mid second trimester, when amnioncentesis is typically performed, finRBCs contain primarily γ-globin with some adult β-globin. Maternal cells contain almost exclusively α and β-globin, with traces of γ-detectable in some samples. Therefore, by measuring the relative expression of the ε, γ and β genes in RBCs purified from maternal blood samples, the presence of fetal cells in the sample can be determined. Furthermore, positive controls can be utilized to assess failure of the FISH analysis itself.

In various embodiments, fetal cells are distinguished from maternal cells based on the differential expression of hemoglobins β, γ or ε. Expression levels or RNA levels can be determined in the cytoplasm or in the nucleus of cells. Thus in some embodiments, the methods herein involve determining levels of messenger RNA (mRNA), ribosomal RNA (rRNA), or nuclear RNA (nRNA).

In some embodiments, identification of finRBCs can be achieved by measuring the levels of at least two hemoglobins in the cytoplasm or nucleus of a cell. In various embodiments, identification and assay is from 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 or 20 fetal nuclei. Furthermore, total nuclei arrayed on one or more slides can number from about 100, 200, 300, 400, 500, 700, 800, 900, 10,000, 100,000, 1,000,000, 2,000,000 to about 3,000,000. In some embodiments, a ratio for γ/β or ε/β is used to determine the presence of fetal cells, where a number less than one indicates that a finRBC(s) is not present. In some embodiments, the relative expression of ε/β or ε/β provides a finRBC index (“FINI”), as measured by γ or ε relative to β. In some embodiments, a FINI for γ/β greater than 5, 10, 15, 20, 25, 30, 35, 40, 45, 90, 180, 360, 720, 975, 1020, 1024, 1250 to about 1250, indicate that a finRBC(s) is present. In yet other embodiments, a FINI for γ/β of less than about 1 indicates that a finRBC(s) is not present. Preferably, the above FINI is determined from a sample obtained during a first trimester. However, similar ratios can be used during second trimester and third trimester.

In some embodiments, the expression levels are determined by measuring nuclear RNA transcripts including, nascent or unprocessed transcripts. In another embodiment, expression levels are determined by measuring mRNA, including ribosomal RNA. There are many methods known in the art for imaging (e.g., measuring) nucleic acids or RNA including, but not limited to, using expression arrays from Affymetrix, Inc. or Illumina, Inc.

RT-PCR primers can be designed by targeting the globin variable regions, selecting the amplicon size, and adjusting the primers annealing temperature to achieve equal PCR amplification efficiency. Thus TaqMan probes can be designed for each of the amplicons with well-separated fluorescent dyes, Alexa fluor®-355 for ε, Alexa Fluor®-488 for γ, and Alexa Fluor®-555 for β. The specificity of these primers can be first verified using ε, γ, and β cDNA as templates. The primer sets that give the best specificity can be selected for further assay development. As an alternative, the primers can be selected from two exons spanning an intron sequence to amplify only the mRNA to eliminate the genomic DNA contamination.

The primers selected can be tested first in a duplex format to verify their specificity, limit of detection, and amplification efficiency using target cDNA templates. The best combinations of primers can be further tested in a triplex format for its amplification efficiency, detection dynamic range, and limit of detection.

Various commercially available reagents are available for RT-PCR, such as One-step RT-PCR reagents, including Qiagen One-Step RT-PCR Kit and Applied Biosystem’s TaqMan One-Step RT-PCR Master Mix Reagents Kit. Such reagents can be used to establish the expression ratio of ε, γ, and β using purified RNA from enriched samples. Forward primers can be labeled for each of the targets, using Alexa fluor®-355 for ε, Alexa fluor®-488 for γ, and Alexa fluor®-555 for β. Enriched cells can be deposited by cytopinning onto glass slides. Additionally, cytopinning the enriched cells can be performed after in situ RT-PCR. Thereafter, the presence of the fluorescent-labeled amplicons can be visualized by fluorescence microscopy. The reverse transcription time and PCR cycles can be optimized to maximize the amplicon signal:background ratio to have maximal separation of fetal over maternal signature. Preferably, signal:background ratio is greater than 5, 10, 50 or 100 a the overall cell loss during the process is less than 50, 10 or 5%.

Fetal Cell Analysis

In step 125, DNA is extracted and purified from cells/nuclei of the enriched product (mixed sample enriched) and reference sample. Methods for extracting DNA are known to those skilled in the art.

In step 131, the DNA is optionally pre-amplified to increase the overall quantity of DNA for subsequent analysis. Pre-amplification of DNA can be conducted using any amplification method known in the art, including for example, amplification via multiple displacement amplification (MDA) (Gonzalez J M, et al. Cold Spring Harb Symp Quant Biol; 68:69-78 (2003), Murthy et al. Hum Mutat 26(2):145-52 (2005) and Paulaud et al., Biotechniques; 38(4):533-4, 556, 558-9 (2005)), and linear amplification methods such as in vitro transcription (Liu, et al., BMC Genomics; 4(1)19 (2003)).
Other methods for pre-amplification include PCR methods including quantitative PCR, quantitative fluorescent PCR (QF-PCR), multiplex fluorescent PCR (MF-PCR), real time PCR (RT-PCR), single cell PCR, PCR-RFLP/RT-PCR-RFLP, hot start PCR and Nested PCR. For example, the PCR products can be directly sequenced bi-directionally by dye-terminator sequencing. PCR can be performed in a 384-well plate in a volume of 15 ul containing 5 ng genomic DNA, 2 mM MgCl2, 0.75 mM DMSO, 1 M Betaine, 0.2 mM dNTPs, 20 pmol primers, 0.2 ul AmpliTaq Gold® (Applied Biosystems), IX buffer (supplied with AmpliTaq Gold). Thermal cycling conditions are as follows: 95°C for 10 minutes; 95°C for 30 seconds, 60°C for 30 seconds, 72°C for 1 minute for 30 cycles; and 72°C for 10 minutes. PCR products can be purified with Ampure® Magnetic Beads (Agencourt) and can be optionally separated by capillary electrophoresis on an ABI3730 DNA Analyzer (Applied Biosystems).

Suitable amplification methods include the ligation chain reaction (LCR), transcription amplification, self-sustained sequence replication, selective amplification of target polynucleotide sequences, consensus sequence primed polymerase chain reaction (CP-PCR), arbitrarily primed polymerase chain reaction (AP-PCR) and nucleic acid based sequence amplification (NASBA). Other amplification methods that may be used in step 131 include those described in U.S. Pat. Nos. 5,242,794, 5,494,810, 4,988,617 and 6,582,938, each of which is incorporated herein by reference.

The pre-amplification step increases the amount of enriched fetal DNA thus allowing analysis to be performed even if up to 1 µg, 500 ng, 200 ng, 100 ng, 50 ng, 40 ng, 30 ng, 20 ng, 10 ng, 5 ng, 1 ng, 500 pg, 200 pg, 100 pg, 50 pg, 40 pg, 30 pg, 20 pg, 10 pg, 5 pg, or 1 pg of fetal or total DNA was obtained from the mixed sample, or between 1-5 µg, 5-10 µg, 10-50 µg of fetal or total DNA was obtained from the mixed sample.


In some embodiments, cDNAs, which are reverse transcribed from mRNAs obtained from fetal or maternal cells, are analyzed for the presence of SNPs using the methods disclosed within. The type and abundance of the cDNAs can be used to determine whether a cell is a fetal cell (such as by the presence of Y chromosome specific transcripts) or whether the fetal cell has a genetic abnormality (such as aneuploidy, abundance of alternative transcripts or problems with DNA methylation or imprinting).

In one embodiment, fetal or maternal cells or nuclei are enriched using one or more methods disclosed herein. Preferably, fetal cells are enriched by flowing the sample through an array of obstacles that selectively directs particles or cells of different hydrodynamic sizes into different outlets such that fetal cells and cells larger than fetal cells are directed into a first outlet and one or more cells or particles smaller than the rare cells are directed into a second outlet.

Total RNA or poly-A mRNA is then obtained from enriched cell(s) (fetal or maternal cells) using purification techniques known in the art. Generally, about 1 µg-2 µg of total RNA is sufficient. Next, a first-strand complementary DNA (cDNA) is synthesized using reverse transcriptase and a single T7-oligo(dT) primer. Next, a second-strand cDNA is synthesized using DNA ligase, DNA polymerase, and RNase enzyme. Next, the double stranded cDNA (ds-cDNA) is purified.

In another embodiment, total RNA is extracted from enriched cells (fetal cells or maternal cells). Next, a two-one-quarter scale Message Amp II reactions (Ambion, Austin, Tex.) are performed for each RNA extraction using 200 ng of total RNA. MessageAmp is a procedure based on antisense RNA (aRNA) amplification, and involves a series of enzymatic reactions resulting in linear amplification of exceedingly small amounts of RNA for use in array analysis. Unlike exponential RNA amplification methods, such as NASBA and RT-PCR, aRNA amplification maintains representation of the starting mRNA population. The procedure begins with total or poly(A) RNA that is reverse transcribed using a primer containing both oligo(dT) and a T7 RNA polymerase promoter sequence. After first-strand synthesis, the reaction is treated with RNase H to cleave the mRNA into small fragments. These small RNA fragments serve as primers during a second-strand synthesis reaction that produces a double-stranded cDNA template.

Any DNA microarray that is capable of detecting one or more SNPs can be used with the methods herein. DNA microarrays comprise a plurality of genetic probes immobilized at discrete sites (i.e., defined locations or assigned positions) on a substrate surface. A DNA microarray preferably monitors at least 5, 10, 20, 50, 100, 200, 500, 1,000, 2,000, 5,000, 10,000, 20,000, 50,000, 100,000, 200, 000 or 500,000 different SNPs. Such SNPs can be located in one or more target chromosomes or on the entire genome. Methods for manufacturing DNA microarrays for detecting SNPs are known in the art. Microarrays that can be used in the systems herein include those commercially available from Affymetrix (Santa Clara, Calif.), Illumina (San Diego, Calif.), Spectral Genomics, Inc. (Houston, Tex.), and Vysis Corporation (Downers Grove, Ill.). Methods for detecting SNPs using microarrays are further described in U.S. Pat. Nos. 6,300,063, 5,837,832, 6,969,589, 6,040,138, and 6,858,412.

In one embodiment, SNPs are detected using molecular inversion probes (MIPs). MIPs are nearly circu-
larized probes having a first end of the probe complementary to a region immediately upstream of the SNP to be detected, and a second end of the probe complementary to a region immediately downstream of the SNP. To use MIPs both ends are allowed to hybridize to genomic regions surrounding the SNP and an enzymatic reaction fills the gap at the SNP position in an allele specific manner. The fully circular probe now can be separated by a simple exonuclease reaction which leaves a primer sequence coupled to a label unique to the allele. The primer is subsequently used to amplify the label which is then hybridized to an array for detection.

[0092] FIG. 5 illustrates one embodiment of an allele specific extension and ligation reaction. Genomic DNA fragments are first annealed to a solid support. Subsequently, probes designed to be unique for each allele (P1 and P2) are annealed to the target DNA. After a washing step, allele-specific primer extension is conducted to extend the probes if such probes have 3' ends that are complementary to their cognate SNP in the genomic DNA template. The extension is followed by ligation of the extended templates to their corresponding locus-specific probes (P3) to create PCR templates. Requiring the joining of two fragments (P1 and P3 or P2 and P3) to create a PCR template provides an additional level of genomic specificity, because any residual incorrectly hybridized allele-specific or locus-specific probes are unlikely to be adjacent and thus should not be able to ligate. Next, fluorescently labeled primers, each with a different dye, are added for PCR amplification, thus providing a means for detection and quantification of each SNP by providing data points. In addition, each SNP is assigned a different address sequence (P3) which is contained within the locus-specific probe. Each address sequence is complementary to a unique capture sequence that can be contained by one of several bead types present in an array. Furthermore, the use of universal PCR primers to associate a fluorescent dye with each SNP allele provides a cost-saving element, because only three primers, two labeled and one unlabeled, are needed regardless of the number of SNPs to be assayed.

[0093] If the addresses are captured by beads, multiple SNPs can be amplified in the same or in different reactions using bead amplification. When more than one DNA polymorphism is used in the same amplification reaction, primers are chosen to be multiplexable (fairly uniform melting temperature, absence of cross-priming on the human genome, and absence of primer-primer interaction based on sequence analysis) with other pairs of primers. Furthermore, primers and loci may be chosen so that the amplicon lengths from a given locus do not overlap with those from another locus. Multiple dyes and multi-color fluorescence readout may be used to increase the multiplexing capacity.

[0094] In some embodiments, highly parallel SNP detection is performed by arrayed primer extension (APEX). In order to perform APEX, a gene locus is chosen where one wishes to analyze SNPs or mutations, for example, loci for abnormal ploidy disorders (e.g. chromosome X, 13, 18, and 21). Oligonucleotides (e.g., about 20-, 25-, 30-, 40-, 50-mers) are designed to be complementary to the gene up to, but not including the base where the mutation or SNP exists. In one example, the oligonucleotides are modified with an amine group at the 5' end to facilitate covalent binding to activated glass slides, in this case epoxy silanized surfaces. The locus in question is PCR amplified and the DNA enzymatically sheared to facilitate hybridization to the oligos. The PCR reactions contain dTTP and dUTP at about a 5 to 1 ratio, and the incorporation of the dUTP allows the amplified DNA to be enzymatically cut with uracil N-glycosylase (UNG). The optimal size of the sheared DNA is about 100 base pairs. The sheared DNA is then hybridized to the bound oligos and a primer extension reaction carried out using a thermostable DNA polymerase such as Thermo Sequenase (Amersham Pharmacia Biotech) or AmpliTaq FS (Roche Molecular Systems). The primer extension reaction contains four deoxynucleotides (ddNTPs) corresponding to A, C, G & T, with each ddNTP containing a distinct fluor molecule. In the above example, ddNTPs can be conjugated to either fluorescein, Cy3™, Texas Red or Cy5™. Depending on which base is next in the sequence (wild type, mutant or SNP), the primer extension reaction will incorporate one nucleotide with one and only one of the four dyes. Thus, by applying a simple four laser scan one can tell which base is next in the sequence as each of the above dyes are easily spectrally separable. A large number of different oligos, (e.g., 5-, 10-, 15-, 20-, 30-, 40-, 50-, 60-, 70-, 80-, 90-, 100-thousand probes) may be attached to a slide for this type of analysis with the requirement that very little cross hybridization occurs among all the sequences. It may be helpful to increase the length of the oligos (e.g., 50-, 60-, 70-, 80-mers) so that the initial hybridization can be done at higher stringency resulting in less background from non-homologous hybridization. In the APEX method, the signal to noise ratio is about 40 to 1, a level which is more than sufficient for unambiguously identifying SNPs and mutations. To design such large arrays for SNP analysis, a computational screen can be conducted to favor a subset of sequences with similar GC content and thermodynamic properties, and eliminate sequences with possible secondary structure or sequence similarity to other tags, Shoemaker et al. Nature Genetics 14:450-456 (1996); Giaever et al. Nature Genetics 21:278-283 (1999); Winzeek et al. Science 285:901-906 (1999). For example, in high density tag array 64,000 probes, each probe occupying an area of 30x30 μm, are used for parallel genotyping of human SNPs.

[0095] In some cases, it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. Gasparini et al., Mol. Cell Probes 6:1 (1992). Amplification is subsequently performed using Tag ligase and the like. Barany, Proc. Natl. Acad. Sci. USA 88:189 (1991). In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

[0096] Alternatively, detection of single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (e.g., SNP), Orita et al., Proc. Natl. Acad. Sci. USA: 86: 2766 (1989); Cotton, Mutat. Res.: 285: 125-144 (1993); and Hayashi, Genet. Anal. Tech. Appl.: 9: 73-79 (1992). Single-stranded DNA fragments of sample and control nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in
which the secondary structure is more sensitive to a change in sequence. The subject method utilizes heteroduplex analysis to separate double-stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. Keen et al., Trends Genet. 7: 5 (1991).

[0097] Other methods for detecting SNPs include methods in which protection from cleavage agents is used to detect mismatched bases in DNA/RNA or RNA/DNA heteroduplexes. Myers et al., Science 230: 1242 (1985). In general, the art technique of “mismatch cleavage” starts by providing heteroduplexes of formed by hybridizing labeled DNA or RNA containing the control sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single stranded regions of the duplexes, such as those that exist due to “base pair mismatches” between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digest the mismatched regions. Furthermore, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. Cotton et al., Proc. Natl. Acad. Sci. USA 85:4397 (1988); and Saleeba et al., Methods Enzymol. 217: 286-295 (1992). The control DNA or RNA can be labeled for detection.

[0098] SNPs can also be detected and quantified using by sequencing methods including the classic Sanger sequencing method as well as high throughput sequencing, which may be capable of generating at least 1,000, 5,000, 10,000, 20,000, 30,000, 40,000, 50,000, 100,000 or 500,000 sequence reads per hour, with at least 50, 60, 70, 80, 90, 100, 120 or 150 bases per read.

[0099] High throughput sequencing can involve sequencing-by-synthesis, sequencing-by-ligation, and ultra deep sequencing.

[0100] Sequence-by-synthesis can be initiated using sequencing primers complementary to the sequencing element on the nucleic acid tags. The method involves detecting the identity of each nucleotide immediately after (substantially real-time) or upon (real-time) the incorporation of a labeled nucleotide or nucleotide analog into a growing strand of a complementary nucleic acid sequence in a polymerase reaction. After the successful incorporation of a label nucleotide, a signal is measured and then nullified by methods known in the art. Examples of sequence-by-synthesis methods are described in U.S. Application Publication Nos. 2003/0044781, 2006/0024711, 2006/0024678 and 2005/0100932. Examples of labels that can be used to label nucleotide or nucleotide analogs for sequencing-by-synthesis include, but are not limited to, chromophores, fluorescent moieties, enzymes, antigens, heavy metal, magnetic probes, dyes, phosphorescent groups, radioactive materials, chemiluminescent moieties, scattering or fluorescent nanoparticles, Raman signal generating moieties, and electrochemical detection moieties. Sequencing-by-synthesis can generate at least 1,000, at least 5,000, at least 10,000, at least 20,000, 30,000, at least 40,000, at least 50,000, at least 100,000 or at least 500,000 reads per hour. Such reads can have at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 120 or at least 150 bases per read.

[0101] Another sequencing method involves hybridizing the amplified regions to a primer complementary to the sequence element in an LST. This hybridization complex is incubated with a polymerase, ATP sulfurylase, luciferase, apyrase, and the substrates luciferin and adenosine 5’ phosphosulfate. Next, deoxynucleotide triphosphates corresponding to the bases A, G, C, T (U) are added sequentially. Each base incorporation is accompanied by release of pyrophosphate, converted to ATP by sulfurylase, which drives synthesis of oxyluciferin and the release of visible light. Since pyrophosphate release is equimolar with the number of incorporated bases, the light given off is proportional to the number of nucleotides adding in any one step. The process is repeated until the entire sequence is determined.

[0102] Yet another sequencing method involves a four-color sequencing by ligation scheme (degenerate ligation), which involves hybridizing an anchor primer to one of four positions. Then an enzymatic ligation reaction of the anchor primer to a population of degenerate nonamers that are labeled with fluorescent dyes is performed. At any given cycle, the population of nonamers that is used is structure such that the identity of one of its positions is correlated with the identity of the fluorophores attached to that nonamer. To the extent that the ligase discriminates for complementarily at that queried position, the fluorescent signal allows the inference of the identity of the base. After performing the ligation and four-color imaging, the anchor primer:nonamer complexes are stripped and a new cycle begins. Methods to image sequence information after performing ligation are known in the art.

[0103] In some cases, high throughput sequencing involves the use of ultra deep sequencing, such as described in Marguiles et al., Nature 437 (7057): 376-80 (2005). Briefly, the amplicons are diluted and mixed with beads such that each bead captures a single molecule of the amplified material. The DNA molecule on each bead is then amplified to generate millions of copies of the sequence which all remain bound to the bead. Such amplification can occur by PCR. Each bead can be placed in a separate well, which can be a (optionally addressable) picolitre-sized well. In some embodiments, each bead is captured within a droplet of a PCR-reaction-mixture-in-oil-emulsion and PCR amplification occurs within each droplet. The amplification on the bead results in each bead carrying at least 1 million, at least 5 million, or at least 10 million copies of the original amplicon coupled to it. Finally, the beads are placed into a highly parallel sequencing by synthesis machine which generates over 400,000 reads (~100 bp per read) in a single 4 hour run.


[0105] The microarray or sequencing methods described herein provide a readout, which can be visualized via apparatus and methods known in the art. For example, for a given marker or at a given tag probe position, the fluorescence intensity of each of the fluorophores utilized (e.g., tagged sequencing or PCR primers) provides a signal which is detected by apparatus or automated systems/machines
known in the art. The fluorophore markers can be utilized either in an array-based or sequencing-based analysis.

[0106] In step 151, SNP data is used to determine aneuploidy by, e.g., determining the ratio of material allele(s) to parental allele(s) (or vice versa); or determining ratio of maternal allele(s) in a region suspected of aneuploidy versus in a control region.

[0107] Aneuploidy means the condition of having less than or more than the normal diploid number of chromosomes. In other words, it is any deviation from euploidy. Aneuploidy includes conditions such as monosomy (the presence of only one chromosome of a pair in a cell’s nucleus), trisomy (having three chromosomes of a particular type in a cell’s nucleus), tetrasomy (having four chromosomes of a particular type in a cell’s nucleus), pentasomy (having five chromosomes of a particular type in a cell’s nucleus), triploidy (having three of every chromosome in a cell’s nucleus), and tetraploidy (having four of every chromosome in a cell’s nucleus). Birth of a live tripliod is extraordinarily rare and such individuals are quite abnormal; however triploidy occurs in about 2-3% of all human pregnancies and appears to be a factor in about 15% of all miscarriages. Tetraploidy occurs in approximately 9% of all miscarriages. (http://www.emedicine.com/med/topic3241.htm).

[0108] In one embodiment, kits are provided which include a separation device, optionally a capture device and the reagents and devices used for the analysis of the genomic sequences. For example, the kit may include the separation arrays and DNA microarrays for detecting one or more SNPs. Any of the devices mentioned for the DNA determination may be combined with the separation devices. The combination of the array separation devices with DNA analysis devices provides gentle handling and accurate analysis.

[0109] A simple intuitive understanding of the effect of trisomy is that it increase the abundances of fetal alleles at loci within the affected region. Trisomies are predominately from maternal non-disjunction events, so typically both maternal alleles, and a single paternal allele, are increased, and the ratio of maternal allele abundance to paternal allele abundance is higher in the trisomic region. These signatures may be masked by differences in DNA amplification and hybridization efficiency from locus to locus, and from allele to allele.

[0110] In one embodiment, trisomies are determined by comparing abundance (e.g., intensities) of maternal and paternal alleles in a genomic region. Within a locus, the PCR differences are smaller than between loci, because the same primers are responsible for all the different allele amplicons at that locus. Therefore, the allele ratios may be more stable than the overall allele abundances. This can be exploited by identifying loci where the paternal allele is distinct from the maternal allele and taking the ratio of the paternal allele strength to the average of the maternal allele strengths. These allele ratios then can be averaged over the hypothesized aneuploidy region and compared to the average over a control region. The distributions of these ratio values in the hypothesized aneuploidy region and in the control region can be compared to create an estimate of statistical significance for the observed difference in means. A simple example of this procedure would use Student's t-test.

[0111] Thus, the present invention contemplates detection of fetal abnormality by determining a ratio of abundance of maternal allele(s) and abundance of paternal allele(s) (or vice versa) in one or more genomic regions of interest. (Preferably the paternal allele differs from one or both the maternal alleles). The genomic region can be derived from a mixed sample comprising fetal and maternal cells. The sample can be obtained from a pregnant animal and can be, e.g., a blood sample. In some cases, the genomic region includes a SNP and/or an informative SNP. In some cases at least 10, 20, 50, 100, 200 or 500 SNPs are analyzed per sample. The SNPs analyzed can be in a single locus, different loci, single chromosome, or different chromosomes. In some cases, a first genomic region (SNP) analyzed is in a genomic region suspected of being trisomic or is trisomic and a second genomic region (SNP) analyzed is in a control region that is non-trisomic or a region suspected of being non-trisomic. The ratio of alleles (e.g., maternal/paternal) in a first genomic region or first plurality of genomic regions (trisomic) (hereinafter test regions) is then compared with a ratio of alleles (e.g., maternal/paternal) in the second genomic region or second plurality of genomic regions (hereinafter control regions). The control region(s) and test region(s) can be on the same or different chromosomes. In some instances, comparison is made by determining the difference in means of the ratios in the first regions and second regions. Detection of an increase of paternal abundance in the test region(s) is indicative of paternal trisomy, while detection of an increase of maternal abundance in the first regions is indicative of paternal trisomy. Furthermore, calculation of error rate based on amplification can be performed prior to making a call if a fetus has a specific condition (e.g., trisomy) or not.

[0112] Alternatively, the maternal allele strengths over the suspected aneuploid region(s) can be compared to those in the control region(s), all without forming ratios to paternal alleles. In this approach, errors in the measurement of the paternal allele abundances are not calculated. However, differences in amplification efficiency between primer pairs are calculated. These measurements can be larger than differences between alleles in the same locus. In this approach there also may be a residual bias between the efficiencies averaged over certain chromosomes. Therefore it may be useful to also perform the same detection process in a reference sample (e.g. maternal only cell sample) and then take the ratio of ratios. In other words, the ratio obtained for the mixed sample of the abundance in test genomic region(s) and control genomic region(s) divided by the same ratio obtained from the reference sample. The ratios obtained for the mixed and reference samples reflect allele strength over suspected aneuploidy region over allele strength over control region, and the ratio of ratios presents an estimate that is normalized to the reference (maternal) sample. Such ratio of ratios is therefore free of chromosome bias, but may include errors in the measurements of the reference sample, as that sample is used as the control or normalizer.

[0113] In some cases, the methods herein contemplate detecting fetal abnormality by comparing an abundance of one or more maternal allele(s) in a first genomic region or regions (test region(s)) with one or more maternal alleles in a second genomic region or regions (control region(s)) in a mixed sample (e.g., maternal blood sample from pregnant animal). This ratio can then be compared to a similar ratio.
Aneuploidy can be determined by modeling SNP data. One example of a model for SNP data in the context of fetal diagnosis is given in Equations 1-3 below:

\[ s_{k} = \delta_{k} \left[ 1 - f \left( m_{1} + m_{3} \right) \right] \left( m_{1} \text{ or } m_{3} \right) p_{s} \text{ residual} \quad (1) \]

A trisomy caused by maternal non-dysjunction is represented by

\[ s_{k} = \delta_{k} \left[ 1 - f \left( m_{1} + m_{3} \right) \right] \left( m_{1} \text{ or } m_{3} \right) p_{s} \text{ residual} \quad (2) \]

And a paternally inherited trisomy is represented by

\[ s_{k} = \delta_{k} \left[ 1 - f \left( m_{1} + m_{3} \right) \right] \left( m_{1} \text{ or } m_{3} \right) p_{s} \text{ residual} \quad (3) \]

In Equations 1-3, \( \delta_{k} \) denotes a scale factor which subsumes the efficiencies of amplification, hybridization, and readout common to the alleles at locus \( k \). In this model amplification differences between different primer pairs are fitted and do not appear in the residuals. Alternatively, a single \( \delta_{k} \) parameter could be used and the residuals would reflect these differences. Further, \( f \) represents the fraction of fetal cells in the mixture, \( m_{1} \) and \( m_{3} \) denote the maternal alleles at locus \( k \), and \( p_{s} \) denotes the paternal allele at locus \( k \). The allele symbols actually represent unit data contributions that can be arithmetically summed; e.g., \( m_{3} \) might be a detection of the 'C' genotype represented by unit contribution to the 'C' bin at that locus.

Equations 1-3 represent five different models because of the ambiguity between \( m_{1} \) and \( m_{3} \) in the last equations. In other words, since Equation 1 and 3 are different and in each equation there are two possibilities (i.e., \( m_{1} \) or \( m_{3} \)) then it follows that each of Equations 1 and 3 represent two different models. Therefore, Equations 1-3 represent five different models. Testing for aneuploidy of Chromosomes 13, 18, and 21, for example entails 5x5x5x125 different model variants that would be fit to the data.

The parameter values for the maternal allele identities are taken from the results for the reference (i.e., maternal-only) sample and the remaining parameters are fit to the data from the mixed sample. Because the number of parameters is very large when the number of loci is large, a global optimization requires iterative search techniques. One possible approach is to do the following for each model variant:

1. Set \( f \) to 0 and solve for \( A_{k} \) at each locus.
2. Set \( f \) to a value equal to the smallest fetal/maternal cell ratio for which fetal cells are likely to be detectable.
3. Solve for paternal allele(s) identities and strengths at each locus, one locus at a time, that minimize data-model residuals.
4. Fix the paternal alleles and adjust \( f \) to minimize residuals over all the data.
v) Now vary only the $A_k$ to minimize residuals. Repeat iv and v until convergence.

vi) Repeat iii through v until convergence.

The best overall fit of model to data is selected from among all the model variants. The best overall fit yields the values of $f$ and $A_k$ we will call $f_{\text{max}}$, $A_{\text{max}}$. The likelihood of observing the data given $f_{\text{max}}$, $A_{\text{max}}$ can be compared to the likelihood given $f=0$. The ratio is a measure of the amount of evidence for fetal DNA. A typical threshold for declaring fetal DNA would be a likelihood ratio of ~1000 or more. The likelihood calculation can be approximated by a more familiar Chi-squared calculation involving the sum of squared residuals between the data and the model, where each residual is normalized by the expected rms error. This Chi-squared is a good approximation to the Log(likelihood) to the extent the expected errors in the data are Gaussian additive errors, or can be made so by some amplitude transformation of the data.

If based on the above determination of likelihood ratio it is decided that fetal DNA is not present, then the test is declared to be non-informative. If it is decided that fetal DNA is present, then the likelihoods of the data given the different data model types can be compared to declare aneuploidy. The likelihood ratios of aneuploidy models (Equations 2 and 3) to the normal model (Equation 1) are calculated and these ratios are compared to a predefined threshold. Typically this threshold is set so that in controlled tests all the trisomic cases are declared aneuploid. Thus, it is expected that the vast majority (>99.9%) of all truly trisomic cases are declared aneuploid by the test. Another approach to accomplish approximately the 99.9% detection rate is to increase the likelihood ratio threshold beyond that necessary to declare all the known trisomic cases in the validation set by a factor of 1000/N, where N is the number of trisomy cases in the validation set.

In step 161 (FIG. 1), which is optional, the presence of fetal cells and ratio of fetal alleles/maternal alleles is determined. Because the fraction of fetal cells can be small or even zero, the aneuploidy signal (the departure of the observed ratio from unity) may be weak even when fetal aneuploidy is present. An independent estimate of fetal cell fraction, including a confidence estimate of whether measurable fetal DNA is present at all, is useful in interpreting the observed aneuploidy ratios. FIG. 7 illustrates allele signals re-ordered by rank. Assuming the mother has no more than two alleles at each locus, the magnitude of the third ranked allele is potentially a robust indicator of the presence of fetal DNA. Although measurement errors can artificially inflate the size of the third and fourth alleles, it is very unlikely to result in a bimodal distribution for the relative magnitude of the third allele with respect to the first two. Such a bimodal distribution is illustrated in FIG. 8. The secondary peak of this distribution occurs at a value approximately equal to the fraction of fetal cells. (This is one way to determine the value of the variable f in the data model.) The statistical confidence that the bimodality is real can be used to assign a confidence that fetal DNA was present in the mixed sample. Statistical tests for bimodality are discussed in M. Y. Cheng and P. Hall, J.R. Statist. Soc. B (1998), 60 (Part 3) pp. 579-587. If this confidence level exceeds a threshold, e.g., 90%, 95%, 99% or 99.9%, an aneuploidy call may be made: The threshold set can be stringent (e.g., 99.9%)

to avoid declaring a fetus normal when in fact it is not. Thus, the independently estimated fetal cell fraction can be used to interpret the aneuploidy statistic. For example, a value f=0.5 along with an estimated aneuploidy ratio from the fetal-maternal mixture of 0.05±0.01 would weaken the evidence for aneuploidy because the ratio is too small to be consistent with the independently determined f value (the ratio should be ~1+f/2). As another example, a value f=0.1 along with an estimated aneuploidy ratio from the fetal-maternal mixture of 0.05±0.02 would tend to strengthen the evidence for aneuploidy because the observed ratio is consistent with the independently derived value of f.

In any of the embodiments herein SNP data may be analyzed for possible errors. For example, in some instances SNP data can contain small additive errors associated with the readout technology, multiplicative errors associated with DNA amplification and hybridization efficiencies being different from locus to locus and from allele to allele within a locus, and errors associated with imperfect specificity in the process. By including the many parameters ($A_k$) in the model, rather than a single scale parameter, the residuals include allele-to-allele efficiency differences but not locus to locus differences. These tend to be multiplicative errors in the resulting observed allele strengths heights (e.g., two signals may be 20% different in strength although the starting concentrations of the alleles are identical). In other words, by providing many parameters, the errors that are otherwise attributable to locus to locus differences, are minimized. As a first approximation, one can assume errors are random from allele to allele; errors have relatively small additive measurement noise error components; and larger Poisson and multiplicative error components exist. The magnitudes of these error components can be estimated from repeated processing of identical samples. A Chi-square residuals calculation for any data-model fit then can be supported with these modeled squared errors for any peak height or data bin.

For example, we anticipate a large scale SNP genotyping platform such as the Golden Gate assay by Illumina will provide ~100 SNP loci per chromosome of interest. Measurements of repeated 'normal' pregnancy samples would give ratios of paternal to maternal allele strengths which varied by ~20% due to assay errors. Averaged over the 100 loci in a chromosome, the ratio error would be reduced to 20%/sqrt(100), or ~2%. For an assumed fetal/maternal cell ratio of 0.2 in a sample, the expected observed aneuploidy ratio in the case of a trisomy would be 1.10 with an estimation error of 0.02, yielding a confident (5 sigma) detection of aneuploidy.

Alternatively, when using a single $A$ parameter, the residuals will be larger and will contain a component which is correlated between alleles at the same locus. Calculation of likelihood will need to take this correlation into account.

Another aspect of the invention involves a computer executable logic for determining the presence of fetal cells in a mixed sample and fetal abnormalities and/or conditions in such cells. A computer program product is described comprising a computer usable medium having the computer executable logic (computer software program, including program code) stored therein. Computer executable logic when executed by the processor causes the processor to perform one or more functions described
For example, a computer executable logic can be utilized to automate, process or control sample collection, sample enrichment, pre-amplification, SNP data modeling, estimating fetal/maternal allele ratio, comparing maternal allele intensity from suspected aneuploid region and control region and determining the existence of aneuploidy and the type of aneuploidy if one exists.

For example, the computer executable logic can determine the presence and ratio of fetal cells to maternal cells in a mixed sample. The executable code can also receive data for one or more SNPs, and apply such data to one or more data models. The computer executable logic can then calculate a set of values for each of the data sets associated with each data model; select the data model that best fit the data, model and calculate for any potential errors in the data models; for example, a computer executable logic can determine the ratio of maternal alleles to paternal alleles in one or more SNP locations; and/or the ratio of maternal alleles in a region suspected of aneuploidy and a control region. One example of a data model provides a determination of a fetal abnormality from given data signals of SNPs at two genomic regions. The executable logic can establish the presence or absence of trisomy, and conclude whether the trisomy is paternally derived or if it originated from a maternal non-disjunction event. For example, the program can fit SNP data to the following model, which can provide the diagnosis as follows:

\[ x_{k} = \frac{A_{k}[(1-f)(m_{1} + m_{2})]}{q_{k}(m_{1} + m_{2} + 2p_{k})} + \text{residual} \]  

A trisomy caused by maternal non-dysjunction is represented by:

\[ x_{k} = \frac{A_{k}[(1-f)(m_{1} + m_{2})]}{q_{k}(m_{1} + m_{2} + 2p_{k})} + \text{residual} \]  

A paternally inherited trisomy is represented by:

\[ x_{k} = \frac{A_{k}[(1-f)(m_{1} + m_{2})]}{q_{k}(m_{1} + m_{2} + 2p_{k})} + \text{residual} \]  

In Equations 1-3, \( A_{k} \) denotes a scale factor which subsumes the efficiencies of amplification, hybridization, and readout common to the alleles at locus \( k \). In this model amplification differences between different primer pairs are fitted and do not appear in the residuals. Alternatively, a single \( A \) parameter could be used and the residuals would reflect these differences. Further, \( f \) represents the fraction of fetal cells in the mixture, \( m_{1} \) and \( m_{2} \) denote the maternal alleles at locus \( k \), and \( p_{k} \) denotes the paternal allele at locus \( k \). The allele symbols actually represent unit data contributions that can be arithmetically summed; e.g., \( m_{k1} \) might be a detection of the ‘C’ genotype represented by unit contribution to the ‘C’ bin at that locus.

In some cases, the computer executable logic records data measurements corresponding to readouts (e.g., SNP intensities from DNA microarray or a sequencing machine). Such measurements can be processed by the computer executable logic to determine fetal/maternal allele ratios and provide a call with respect to detection of aneuploidy. Moreover, computer executable logic can control display of such results in print or electronic formats, which an operator can view. Thus, a computer executable logic can include code for receiving data on one or more target DNA polymorphisms (i.e. SNP loci); calculating a set of values for each of the data sets associated with each data model; selecting the data model that best fit the data, wherein the best model will be an indication of the presence of fetal cells in the mixed sample and fetal abnormalities and/or conditions in said cells. The determination of presence of fetal cells in the mixed sample and fetal abnormalities and/or conditions in said cells can be made by the computer executable logic or an user. Therefore, the computer based logic can provide results for estimating fetal/maternal ratios, allele strength and aneuploidy, which can be observed by a technician or operator.

EXAMPLES

Example 1

Separation of Fetal Cord Blood

[0143] FIG. 12A-D shows a schematic of the device used to separate nucleated cells from fetal cord blood.

[0144] Dimensions: 100 mm x 28 mm x 1 mm

[0145] Array design: 3 stages, gap size=18, 12 and 8 μm for the first, second and third stage, respectively.

[0146] Device fabrication: The arrays and channels were fabricated in silicon using standard photolithography and deep silicon reactive etching techniques. The etch depth is 140 μm. Through holes for fluid access are made using KOH wet etching. The silicon substrate was sealed on the etched face to form enclosed fluidic channels using a blood compatible pressure sensitive adhesive (9795, 3M, St Paul, Minn.).

[0147] Device packaging: The device was mechanically mated to a plastic manifold with external fluidic reservoirs to deliver blood and buffer to the device and extract the generated fractions.

[0148] Device operation: An external pressure source was used to apply a pressure of 2.0 PSI to the buffer and blood reservoirs to modulate fluidic delivery and extraction from the packaged device.

[0149] Experimental conditions: Human fetal cord blood was drawn into phosphate buffered saline containing Acid Citrate Dextrose anticoagulants. 1 mL of blood was processed at 3 mL/hr using the device described above at room temperature and within 48 hrs of draw. Nucleated cells from the blood were separated from unenculted cells (red blood cells and platelets), and plasma delivered into a buffer stream of calcium and magnesium-free Dulbecco’s Phosphate Buffered Saline (14190-144, Invitrogen, Carlsbad, Calif.) containing 1% Bovine Serum Albumin (BSA) (A8412-10 ML, Sigma-Aldrich, St Louis, Mo.) and 2 mM EDTA (15575-020, Invitrogen, Carlsbad, Calif.).

[0150] Measurement techniques: Cell smears of the product and waste fractions (FIGS. 8A-8B) were prepared and stained with modified Wright-Giemsa (WG16, Sigma Aldrich, St Louis, Mo.).

[0151] Performance: Fetal nucleated red blood cells were observed in the product fraction (FIG. 8A) and absent from the waste fraction (FIG. 8B).

Example 2

Isolation of Fetal Cells from Maternal Blood

[0152] The device and process described in detail in Example 1 were used in combination with immunomagnetic
affinity enrichment techniques to demonstrate the feasibility of isolating fetal cells from maternal blood.

[0153] Experimental conditions: blood from consenting maternal donors carrying male fetuses was collected into K$_2$EDTA vacutainers (366643, Becton Dickinson, Franklin Lakes, N.J.) immediately following elective termination of pregnancy. The undiluted blood was processed using the device described in Example 1 at room temperature and within 9 hrs of draw. Nucleated cells from the blood were separated from enucleated cells (red blood cells and platelets), and plasma delivered into a buffer stream of calcium and magnesium-free Dulbecco’s Phosphate Buffered Saline (14190-144, Invitrogen, Carlsbad, Calif.) containing 1% Bovine Serum Albumin (BSA) (A8412-100ML, Sigma-Aldrich, St Louis, Mo.). Subsequently, the nucleated cell fraction was labeled with anti-CD71 microbeads (130-046-201, Miltenyi Biotech Inc., Auburn, Calif.) and enriched using the MiniMACS™ MS column (130-042-201, Miltenyi Biotech Inc., Auburn, Calif.) according to the manufacturer’s specifications. Finally, the CD71-positive fraction was spotted onto glass slides.

[0154] Measurement techniques: Spotted slides were stained using fluorescence in situ hybridization (FISH) techniques according to manufacturer’s specifications using Vysis probes (Abbott Laboratories, Downer’s Grove, Ill.). Samples were stained from the presence of X and Y chromosomes. In one case, a sample prepared from a known Trisomy 21 pregnancy was also stained for chromosome 21.

[0155] Performance: Isolation of fetal cells was confirmed by the reliable presence of male cells in the CD71-positive population prepared from the nucleated cell fractions (FIGS. 10A-10F). In the single abnormal case tested, the trisomy 21 pathology was also identified (FIG. 11).

Example 3

Confirmation of the Presence of Male Fetal Cells in Enriched Samples

[0156] Confirmation of the presence of a male fetal cell in an enriched sample is performed using qPCR with primers specific for DYZ, a marker repeated in high copy number on the Y chromosome. After enrichment of fnRBC by any of the methods described herein, the resulting enriched fnRBC are binned by dividing the sample into 100 PCR wells. Prior to binning, enriched samples may be screened by FISH to determine the presence of any fnRBC containing an aneuploidy of interest. Because of the low number of fnRBC in maternal blood, only a portion of the wells will contain a single fnRBC (the other wells are expected to be negative for fnRBC). The cells are fixed in 2% Paraformaldehyde and stored at 4°C. Cells in each bin are pelleted and resuspended in 5 μL PBS plus 1 μL 20 mg/ml Proteinase K (Sigma #P-2308). Cells are lysed by incubation at 65°C for 60 minutes followed by inactivation of the Proteinase K by incubation for 15 minutes at 95°C. For each reaction, primer sets (DYZ forward primer TCGAGTCATTCATCCATCC; DYZ reverse primer ATGGAATGGCATCAAACGGAA; and DYZ Taqman Probe 6FAM-TGGCTGTCCATTCCA-MGBNFQ, TaqMan Universal PCR master mix. No AmpErase and water are added. The samples are run and analysis is performed on an ABI 7500: 2 minutes at 50°C, 10 minutes 95°C followed by 40 cycles of 95°C (15 seconds) and 60°C (1 minute). Following confirmation of the presence of male fetal cells, further analysis of bins containing fnRBC is performed. Positive bins may be pooled prior to further analysis.

[0157] FIG. 13 shows the results expected from such an experiment. The data in FIG. 13 was collected by the following protocol. Nucleated red blood cells were enriched from cord cell blood of a male fetus by sucrose gradient two Heme Extractions (HE). The cells were fixed in 2% paraformaldehyde and stored at 4°C. Approximately 10x1000 cells were pelleted and resuspended each in 5 μL PBS plus 1 μL 20 mg/ml Proteinase K (Sigma #P-2308). Cells were lysed by incubation at 65°C for 60 minutes followed by a inactivation of the Proteinase K by 15 minute at 95°C. Cells were combined and serially diluted 10-fold in PBS for 100, 10 and 1 cell per μl final concentration were obtained. Six μl of each dilution was assayed in quadruplicate in 96 well format. For each reaction, primer sets (DYZ forward primer TCGAGTCATTCATCCATCC; DYZ reverse primer ATGGAATGGCATCAAACGGAA; and DYZ Taqman Probe 6FAM-TGGCTGTCCATTCCA-MGBNFQ, TaqMan Universal PCR master mix, No AmpErase and water were added to a final volume of 25 μl per reaction. Plates were run and analyzed on an ABI 7500: 2 minutes at 50°C, 10 minutes 95°C followed by 40 cycles of 95°C (15 seconds) and 60°C (1 minute). These results show that detection of a single fnRBC in a bin is possible using this method.

Example 4

Confirmation of the Presence of Fetal Cells in Enriched Samples by STR Analysis

[0158] Maternal blood is processed through a size-based separation module, with or without subsequent MHEM enhancement of fnRBCs. The enhanced sample is then subjected to FISH analysis using probes specific to the aneuploidy of interest (e.g., triploidy 13, triploidy 18, and XXY). Individual positive cells are isolated by “plucking” individual positive cells from the enhanced sample using standard micromanipulation techniques. Using a nested PCR protocol, STR marker sets are amplified and analyzed to confirm that the FISH-positive aneuploid cell(s) are of fetal origin. For this analysis, comparison to the maternal genotype is typical. An example of a potential resulting data set is shown in Table 2. Non-maternal alleles may be proven to be paternal alleles by paternal genotyping or genotyping of known fetal tissue samples. As can be seen, the presence of paternal alleles in the resulting cells, demonstrates that the cell is of fetal origin (cells #1, 2, 9, and 10). Positive cells may be pooled for further analysis to diagnose aneuploidy of the fetus, or may be further analyzed individually.
<table>
<thead>
<tr>
<th>DNA Source</th>
<th>STR locus D14S</th>
<th>STR locus D16S</th>
<th>STR locus D8S</th>
<th>STR locus F13B</th>
<th>STR locus vWA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal alleles</td>
<td>14, 17</td>
<td>11, 12</td>
<td>12, 14</td>
<td>9, 9</td>
<td>16, 17</td>
</tr>
<tr>
<td>Cell #1 alleles</td>
<td>17</td>
<td>8</td>
<td>15</td>
<td></td>
<td>19</td>
</tr>
<tr>
<td>Cell #2 alleles</td>
<td>17</td>
<td>12</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell #3 alleles</td>
<td>17</td>
<td>12</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell #4 alleles</td>
<td>17</td>
<td>12</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell #5 alleles</td>
<td>17</td>
<td>12</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell #6 alleles</td>
<td>17</td>
<td>12</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell #7 alleles</td>
<td>17</td>
<td>12</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell #8 alleles</td>
<td>17</td>
<td>12</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell #9 alleles</td>
<td>17</td>
<td>12</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell #10 alleles</td>
<td>17</td>
<td>12</td>
<td>9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Example 5

**Confirmation of the Presence of Fetal Cells in Enriched Samples by SNP Analysis**

Maternal blood is processed through a size-based separation module, with or without subsequent MHEM enhancement of InRBCs. The enhanced sample is then subjected to FISH analysis using probes specific to the aneuploidy of interest (e.g., triploidy 13, triploidy 18, and XXY). Samples testing positive with FISH analysis are then binned into 96 microtiter wells, each well containing 15 μL of the enhanced sample. Of the 96 wells, 5-10 are expected to contain a single InRBC and each well should contain approximately 1000 nucleated maternal cells (both WBC and mRBC). Cells are pelleted and resuspended in 5 μL PBS plus 1 μL 20 mg/mL Proteinase K (Sigma #P-2308). Cells are lysed by incubation at 65°C for 60 minutes followed by a inactivation of the Proteinase K by 15 minute at 95°C.

Example 6

**Use of Highly Parallel Genotyping and High Throughput Sequencing for Fetal Diagnosis**

Fetal cells or nuclei can be isolated as described in the enrichment section or as described in example 1. The enrichment process described in example 1 may generate a final mixture containing approximately 500 maternal white blood cells (WBCs), approximately 100 maternal nuclear red blood cells (mRBCs), and a minimum of approximately 10 fetal nucleated red blood cells (InRBCs) starting from an initial 20 ml blood sample taken late in the first trimester. In the context of fetal diagnosis, it is very valuable to have a reference sample containing only the mother’s genome. When the diagnosis procedure is based on enriching for circulating fetal cells in the mother’s blood, the reference sample can be created simply by not enriching for fetal cells, and then diluting enough to ensure that ≤1 fetal cell is expected in the sample used as input to the SNP detection process. Alternatively, white blood cells can be selected, for which the circulating fetal fraction is negligible.

**Example 64**

Perform Multiple Displacement Amplification (MDA): Current technologies and protocols for highly parallel SNP detection with DNA microarray readout result in inaccurate calls when there are too few starting DNA copies or when a particular allele represents a small fraction in the population of input DNA molecules. In the methods described herein a ratio-preserving pre-amplification of the DNA, such as multiple displacement amplification, is done to provide enough copies to support accurate SNP detection via primer extension ligation methods described below. This pre-amplification method is chosen to produce as close as possible the same amplification factor for all target regions of the genome.

**Example 65**

Multiple displacement amplification protocols can be performed as described in Gonzalez et al. Environmental Microbiology 7(7) 1024-1028, (2005). Briefly, samples are suspended in 100 ul 10 mM Tris-HCl buffer (pH 7.5). Cells are lysed by adding 100 ul of alkaline lysis solution (400 mM KOH, 100 mM EDTA) and incubating cells for 10 min on ice. Lysed cells are neutralized with 100 μL of neutralization solution (2 ml 1 M HCl and 3 ml 1 M Tris-HCl). Lysed cells are used directly as template in MDA and PCR reactions.

**Example 66**

1 μl template DNA in 9 μl sample buffer (50 mM Tris-HCl, pH 8.2, 0.5 mM EDTA) containing random
hexamers is denatured at 95°C. for 3 min and placed on ice. Buffer (9 µl) containing dNTPs and 1 µl enzyme mix containing F29 DNA polymerase are added to the 10 µl of denatured DNA template-random hexamers solution and incubated at 30°C. for 6 h. A final incubation at 65°C. for 10 min inactivated the F29 DNA polymerase.

[0167] Highly Parallel Genotyping: Highly parallel SNP detection can be used to obtain information about genotype and gene copy numbers at a large number of loci scattered across the genome, in one procedure. Highly parallel SNP genotyping can be performed as described in Fan et al. Cold. Spring Harb Symp Quant Biol; 68: 69-78, (2003). Genomic DNA is immobilized to streptavidin-coated magnetic beads by mixing 20 µl of DNA (100 ng/µl) with 5 µl of photobiotin (0.2 µg/µl) and 15 µl of mineral oil, and incubating at 95°C. for 30 minutes. Trizma base (25 µl of 0.1 M) is added, followed by two extractions with 75 µl of Sec-butanol to remove unreacted photobiotin. The extracted gDNA (20 µl) is then mixed with 34 µl of Paramagnetic Particle A Reagent (MPA; Illumina) and incubated at room temperature for 90 minutes. The immobilized gDNA is then washed twice with DNA wash buffer (WDI) (Illumina) and resuspended at 10 ng/pl in WDI. In each subsequent reaction, 200 ng (10 µl) of DNA is used.

[0168] Assay oligonucleotides are then annealed to the genomic DNA by combining the immobilized DNA (10 µl) with annealing reagent (MAI; Illumina; 30 µl) and SNP-specific oligonucleotides (10 µl containing 25 mM of each oligonucleotide) to a final volume of 50 µl. LSOs are synthesized with a 5' phosphate group to allow ligation. Annealing is carried out by ramping temperature from 70°C to 30°C over ~8 hours, then holding at 30°C until the next processing step.

[0169] After annealing, excess and mis-hybridized oligonucleotides are washed away, and 37 µl of master mix for extension (MMEI; Illumina) is added to the beads. Extension is carried out at room temperature for 15 minutes. After washing, 37 µl of master mix for ligation (MML; Illumina) is added to the extension products, and incubated for 20 minutes at 57°C to allow the extended upstream oligo to ligate to the downstream oligo.

[0170] The extension products are then amplified by PCR. After extension and ligation, the beads are washed with universal buffer 1 (UB 1; Illumina), resuspended in 35 µl of elution buffer (IPI; Illumina) and heated at 95°C. for one minute to release the ligated products. The supernatant is then used in a 60-µl PCR. PCR reactions are thermocycled as follows: 10 seconds at 25°C; 34 cycles of (35 seconds at 95°C, 35 seconds at 56°C, 2 minutes at 72°C); 10 minutes at 72°C; and cooled to 4°C for 5 minutes. The three universal PCR primers (P1, P2, and P3) are labeled with Cy3, Cy5, and biotin, respectively.

[0171] High throughput sequencing: After the SNP-specific ligation-extension reaction, and amplification of the products, readout of the SNP types can be done using high throughput sequencing as described in Margulies et al. Nature 437 376-380, (2005). Briefly, the amplicons are diluted and mixed with beads such that each bead captures a single molecule of the amplified material. The DNA-carrying beads are isolated in separate 100 um aqueous droplets made through the creation of a PCR-reaction-mixture-in-oil emulsion. The DNA molecule on each bead is then amplified to generate millions of copies of the sequence, which all remain bound to the bead. Finally, the beads are placed into a highly parallel sequencing-by-synthesis machine which can generate over 400,000 sequence reads (~100 bp per read) in a single 4 hour run.

[0172] Fetal Diagnosis: The SNP data obtained from the high throughput sequencing is analyzed for fetal diagnosis using the methods described in Example 9.

Example 7

Use of Highly Parallel Genotyping and Bead Arrays for Fetal Diagnosis

[0173] Fetal cells or nuclei can be isolated as described in the enrichment section or as described in example 1. The enrichment process described in example 1 may generate a final mixture containing approximately 500 maternal white blood cells (WBCs), approximately 100 maternal nuclear red blood cells (mRBCs), and a minimum of approximately 10 fetal nucleated red blood cells (inRBCs) starting from an initial 20 ml blood sample taken late in the first trimester. In the context of fetal diagnosis, it is very valuable to have a reference sample containing only the mother’s genotype. When the diagnosis procedure is based on enriching for circulating fetal cells in the mother’s blood, the reference sample can be created simply by not enriching for fetal cells, and then diluting enough to ensure that <=1 fetal cell is expected in the sample used as input to the SNP detection process. Alternatively, white blood cells can be selected, for which the circulating fetal fraction is negligible.

[0174] Perform Linear Amplification of Genomic DNA: Current technologies and protocols for highly parallel SNP detection with DNA microarray readout result in inaccurate calls when there are too few starting DNA copies or when a particular allele represents a small fraction in the population of input DNA molecules. In the methods described herein a ratio-preserving pre-amplification of the DNA, such as linear amplification of genomic DNA, is done to provide enough copies to support accurate SNP detection via primer extension ligation methods described below. This pre-amplification method is chosen to produce as close as possible the same amplification factor for all target regions of the genome.

[0175] Linear amplification protocols can be performed as described in Liu et al. BMC Genomics 4(1) 19-30 (2003). This protocol uses a terminal transferase tailing step and second strand synthesis to incorporate T7 promoters at the ends of the DNA fragments prior to in vitro transcription (IVT). Briefly, genomic DNA can be obtained either by ChIP or by restriction digests. ChIP DNA is fragmented by sonication and isolated using antibody against dimethyl-H3 K4. Restricted genomic DNA is prepared as follows: genomic DNA isolated by bead lysis, phenol/chloroform extraction, and ethanol precipitation, is restricted either with Alu I or with Rsa I (New England BioLabs (NEB)). Digested products then undergo electrophoresis on a 2% agarose gel. Restriction fragments in the 100-700 bp size range are excised from the gel and purified using the QiAquick Gel Extraction Kit (Qiagen).

[0176] Calf intestinal alkaline phosphatase (CIP) (NEB) is used to reduce 3' phosphate groups from DNA samples prior to IVT. Up to 500 ng DNA is incubated with 2.5 U
enzyme in a 10 µl volume with the supplied buffer at 37° C. for 1 hour. The reaction was cleaned up with the MinElute Reaction Cleanup Kit (Qiagen) per manufacturer instructions except that the elution volume is increased to 20 µl.

[0177] PolyT tails are generated using terminal transferase (TdT) as follows. Up to 50 ng of CIP-treated template DNA is incubated for 20 minutes at 37° C. in a 10 µl solution containing 20 U TdT (NEB), 0.2 M potassium cacodylate, 25 mM Tris-HCl pH 6.6, 0.25 mg/ml BSA, 0.75 mM CoCl₂, 4.6 µM dTTP and 0.4 µM dTTP. The reaction is halted by the addition of 2 µl of 0.5 M EDTA pH 8.0, and product isolated with the MinElute Reaction Cleanup Kit (Qiagen), increasing the elution volume to 20 µl.

[0178] Second strand synthesis and incorporation of the T7 promoter sequence is carried out as follows: the 20 µl tailing reaction product is mixed with 0.6 µl of 25 mM T7-A18B primer (5'-CATTACGCGGCCGCAATATATACGACTCTAGAGGAGA(A)18 [B], where B refers to C, G or T, 5 µl 10x EcoPol buffer (100 mM Tris-HCl pH 7.5, 50 mM MgCl₂, 75 mM dTT, 23 µM dNTPs, and 20.4 µl nucleosome-free water. In experiments with 10-50 ng starting material, the end primer concentration is kept at 300 nM, while the reaction volume is scaled down to maintain an end concentration of 1 ng/ul starting material. For starting amounts less than 10 ng, the volume is kept at 10 µl. If necessary, volume reduction of the elute from the TdT tailing is performed in a vacuum centrifuge on medium heat. Samples are incubated at 94° C. for 2 minutes to denature, ramped down at ~1 °C/sec to 35° C., held at 35° C. for 2 minutes to anneal, ramped down at ~0.5 °C/sec to 25° C. and held while Klenow enzyme is added (NEB) to an end concentration of 0.2 U/µl. The sample is then incubated at 37° C. for 90 minutes for extension. The reaction is halted by addition of 5 µl 0.5 M EDTA pH 8.0 and product is isolated with the MinElute Reaction Cleanup Kit (Qiagen), increasing the elution volume to 20 µl.

[0179] Prior to in vitro transcription, samples are concentrated in a vacuum centrifuge at medium heat to 8 µl volume. The in vitro transcription is performed with the T7 Megascript Kit (Ambion) per manufacturer’s instructions, except that the 37° C. incubation is increased to 16 hours. The samples are purified with the RNeasy Mini Kit (Qiagen) per manufacturer’s RNA cleanup protocol, except with an additional 500 µl wash with buffer RPE. RNA is quantified by absorbance at 260 nm, and visualized on a denaturing 1.25x MOPS-EDTA-Sodium Acetate gel.

[0180] Highly Parallel Genotyping: Highly parallel SNP detection can be used to obtain information about genotype and gene copy numbers at a large number of loci scattered across the genome, in one procedure. Highly parallel SNP genotyping can be performed as described in Example 6.

[0181] Bead Array: After the SNP-specific ligation-extension reaction, and amplification of the products, readout of the SNP types can be done using bead arrays as described in Shen et al., Mutation Research. 573 70-82, (2005). Double-stranded PCR products are immobilized onto paramagnetic particles by adding 20 µl of Paramagnetic Particle B Reagent (MPB; Illumina) to each 60-µl PCR, and incubated at room temperature for a minimum of 60 minutes. The bound PCR products are washed with universal buffer 2 (UB2; Illumina), and denatured by adding 30 µl of 0.1 N NaOH. After one minute at room temperature, 25 µl of the released ssDNAs is neutralized with 25 µl of hybridization reagent (MH I: Illumina) and hybridized to arrays.

[0182] Arrays are hydrated in UB2 for 3 minutes at room temperature, and then preconditioned in 0.1 N NaOH for 30 seconds. Arrays are returned to the UB2-reagent for at least 1 minute to neutralize the NaOH. The pretreated arrays are exposed to the labeled ssDNA samples described above. Hybridization is conducted under a temperature gradient program from 60° C. to 45° C. over ~12 hours. The hybridization is held at 45° C. until the array is processed. After hybridization, the arrays are first rinsed twice in UB2 and once with IS1 (IS1; Illumina) at room temperature with mild agitation, and then imaged at a resolution of 0.8 microns using a BeadArray Reader (Illumina). PMT settings are optimized for dynamic range, channel balance, and signal-to-noise ratio. Cy3 and Cy5 dyes are excited by lasers emitting at 532 nm and 635 nm, respectively.

[0183] The automatic calling of genotypes is performed by genotyping calling software (GenCall) genotyping software, using a Bayesian model, which compared intensities between probes for allele A and allele B across a large number of samples to create archetypal clustering patterns. These patterns allowed the genotyping data to be assigned membership to clusters using a probabilistic model and allowed assignment of a corresponding GenCall score. For example, data points falling between two clusters are assigned a low probability score of being a member of either cluster and had a correspondingly low GenCall score. The cluster quality can be assessed by evaluating the CSS, a measure of statistical separation between clusters. It is defined as

\[
CSS = \max \left[ \frac{|\theta_{AB} - \theta_{AA}|}{\sigma_{|AB - \sigma_{|AA}|}} \right] \left[ \frac{|\theta_{AB} - \theta_{BB}|}{\sigma_{|AB + \sigma_{|BB}|}} \right]
\]

[0184] Loci with cluster scores around the cutoff of 3.0 are visually evaluated and the training clusters refined by manual intervention. A cutoff value of 3.0 can be chosen for the CSS on the basis of minimizing strand concordance errors. Loci with questionable clusters are scored as unsuccessful and excluded from further analysis.

[0185] Fetal Diagnosis: The SNP data obtained from the bead array assay is analyzed for fetal diagnosis using the methods described in Example 9.

Example 8

Use of Highly Parallel Genotyping and DNA Arrays for Fetal Diagnosis

[0186] Fetal cells or nuclei can be isolated as described in the enrichment section or as described in example 1. The enrichment process described in example 1 may generate a final mixture containing approximately 500 maternal white blood cells (WBCs), approximately 100 maternal red blood cells (mRBCs), and a minimum of approximately 10 fetal red blood cells (fRBCs) starting from an initial 20 ml blood sample taken late in the first trimester. In the context of fetal diagnosis, it is very valuable to have a reference sample containing only the mother’s genotype. When the diagnosis procedure is based on enriching for
circulating fetal cells in the mother’s blood, the reference sample can be created simply by not enriching for fetal cells, and then diluting enough to ensure that <1 fetal cell is expected in the sample used as input to the SNP detection process. Alternatively, white blood cells can be selected, for which the circulating fetal fraction is negligible. 

0187] Perform Multiple Displacement Amplification: Current technologies and protocols for highly parallel SNP detection with DNA microarray readout result in inaccurate calls when there are too few starting DNA copies or when a particular allele represents a small fraction in the population of input DNA molecules. In the methods described herein a ratio-preserving pre-amplification of the DNA, such as multiple displacement amplification, is done to provide enough copies to support accurate SNP detection via primer extension ligation methods described below. This pre-amplification method is chosen to produce as close as possible the same amplification factor for all target regions of the genome. Multiple displacement amplification protocols can be performed as described in Example 6.

0188] Highly Parallel Genotyping: Highly parallel SNP detection can be used to obtain information about genotype and gene copy numbers at a large number of loci scattered across the genome, in one procedure. Highly parallel SNP genotyping can be performed as described in Example 6.

0189] DNA Array: After the SNP-specific ligation-extension reaction, and amplification of the products, readout of the SNP type can be done using DNA arrays as described in Gunderson et al. Nature Genetics 37(5) 549-554, (2005). The array data can be obtained using Illumina’s Sentrix BeadArray matrix. Oligonucleotide probes on the beads are 75 bases in length; 25 bases at the 5’ end are used for decoding and the remaining 50 bases are locus-specific. The oligonucleotides are immobilized on activated beads using a 5’ amino group. The array can contain probes for SNP assays (probe pairs, allele A and allele B).

0190] The amplification products of the SNP-specific ligation-extension reaction are denatured at 95°C for 5 min and exposed it to the Sentrix array matrix, which is nated to a microtiter plate, submerging the fiber bundles in 15 ml of hybridization sample. The entire assembly is incubated for 14-18 h at 48°C with shaking. After hybridization, arrays are washed in 1x hybridization buffer and 20% formamide at 48 1C for 5 min.

0191] Allele Specific Primer Extension (ASPE) can be used to score SNPs. Before carrying out the array-based primer extension reaction, Sentrix array matrices are washed for 1 min with wash buffer (33.3 mM NaCl, 3.3 mM potassium phosphate and 0.1% Tween-20, pH 7.6) and then incubated for 15 min in 50 µl of ASPE reaction buffer (Illumina EMM, containing polymerase, a mix of biotin-labeled and unlabeled nucleotides, single-stranded binding protein, bovine serum albumin and appropriate buffers and salts) at 37°C. After the reaction, the arrays are immediately stripped in freshly prepared 0.1 N NaOH for 2 min and then washed and neutralized twice in 1x hybridization buffer for 30 s. The biotin-labeled nucleotides incorporated during primer extension using a sandwich assay is then detected as described in Pinkel et al. PNAS 83 (1986) 2934-2938. The arrays are blocked at room temperature for 10 min in 1 mg ml⁻¹ bovine serum albumin in 1x hybridization buffer and then washed for 1 min in 1x hybridization buffer. The arrays are then stained with streptavidin-phycoerythrin solution (1x hybridization buffer, 3 µg ml⁻¹ streptavidin-phycoerythrin (Molecular Probes) and 1 mg ml⁻¹ bovine serum albumin) for 10 min at room temperature. The arrays are washed with 1x hybridization buffer for 1 min and then counterstained with an antibody reagent (10 mg ml⁻¹ biotinylated antibody to streptavidin (Vector Labs) in 1x PBS/T (137 mM NaCl, 2.7 mM KCl, 4.3 mM sodium phosphate, 1.4 mM potassium phosphate and 0.1% Tween-20) supplemented with 6 mg ml⁻¹ goat normal serum) for 20 min. After counterstaining, the arrays are washed in 1x hybridization buffer and restained with streptavidin-phycoerythrin solution for 10 min. The arrays are washed one final time in 1x hybridization buffer before imaging them in 1x hybridization buffer on a custom CCD-based BeadArray imaging system. The intensities are extracted intoan intensities using custom image analysis software.

0192] The automatic calling of genotypes is performed by genotype calling software (GenCall) genotyping software as described in example 7.

0193] Fetal Diagnosis: The SNP data obtained from the DNA array assay is analyzed for fetal diagnosis using the methods described in Example 9.

Example 9

Fetal Diagnosis

0194] Results obtained in Example 6, 7, and 8 can be used for fetal diagnosis.

0195] A model for SNP data in the context of fetal diagnosis is given in Equations 1-3. A normal (diploid) fetus will result in data xk at locus k

\[ x_k = \delta_k (m_1 + m_2) \delta \delta (m_1 \text{ or } m_2) (p_1 + p_2) \text{ residual} \]  

(1)

A trisomy caused by maternal non-dysjunction is represented by

\[ x_k = \delta_k (m_1 + m_2) \delta \delta (m_1, m_2) (p_1 + p_2) \text{ residual} \]  

(2)

and a paternally inherited trisomy is represented by

\[ x_k = \delta_k (m_1 + m_2) \delta \delta (m_1, m_2) (p_1, p_2) \text{ residual} \]  

(3)

In Equations 1-3, \( \delta_k \) denotes a scale factor which subsumes the efficiencies of amplification, hybridization, and readout common to the alleles at locus k. In this model, amplification differences between different primer pairs are fitted and do not appear in the residuals. Alternatively, a single A parameter could be used and the residuals would reflect these differences. \( \delta \) represents the fraction of fetal cells in the mixture, \( m_1 \) and \( m_2 \) denote the maternal alleles at locus k, and \( p_1 \) denotes the paternal allele at locus k. The allele symbols actually represent unit data contributions that can be arithmetically summed; e.g., \( m_1 \) might be a detection of the ‘C’ genotype represented by unit contribution to the ‘C’ bin at that locus.

0199] FIG. 6 illustrates the kinds of SNP calls that result under this data model. At Locus 1, the fetal genotype was GC. There is a paternally inherited ‘G’ allele contribution in the mixed sample that results in an increase of G signal above the noise level observed in the maternal-only sample, and a maternally inherited ‘C’ allele contribution that increases the C signal. The effective value of \( \delta \) that has been assumed in these illustrations is 0.0-0.2. At Locus 2, the paternal allele is ‘T’. At Locus 3, the fetus is homozygous
In the third row of FIG. 6, the effect of a fetal trisomy is represented by the dashed red lines, superposed on a normal (diploid) mixed-sample pattern. The trisomy is assumed to include Loci 1 and 2, but not Loci 3 and 4. At Loci 1 and 2 both maternal allele strengths are increased in the mixed sample, as well as the separate paternal allele contribution. At Locus 3, it was assumed that the fetus was ‘GG’ and the paternal allele is the same as the first maternal allele. Note that the ratio between the average of the two maternal alleles and the paternal allele will be slightly greater at Loci 1 and 2 than at Locus 4—this is one indicator of trisomy.

Simple, Suboptimal Detection Methods

[0200] A simple intuitive understanding of the effect of trisomy is that it increases the abundances of fetal alleles at loci within the affected region. Trisomies are predominately from maternal non-disjunction events, so typically both maternal alleles, and a single paternal allele, are increased, and the ratio of maternal allele abundance to paternal allele abundance is higher in the trisomic region. These signatures may be masked by differences in DNA amplification and hybridization efficiency from locus to locus, and from allele to allele.

[0201] Within a locus, the PCR differences are smaller than between loci, because the same primers are responsible for all the different allele amplicons at that locus. Therefore, the allele ratios may be more stable than the overall allele abundances. This can be exploited by identifying loci where the paternal allele is distinct from the maternal allele and taking the ratio of the paternal allele strength to the average of the maternal allele strengths. These allele ratios then can be averaged over the hypothesized aneuploidy region and compared to the average over a control region. The distributions of these ratio values in the hypothesized aneuploidy region and in the control region can be compared to create an estimate of statistical significance for the observed difference in means. A simple example of this procedure would use Student's t-test.

[0202] Alternatively, the maternal allele strengths over the suspected aneuploid region can be compared to those in the control region, all without forming any ratios to paternal alleles. In this approach, errors in the measurement of the paternal allele abundances do not enter; however, the differences in amplification efficiency between primer pairs do enter, and these typically will be larger than differences between alleles in the same locus. In this approach there also may be a residual bias between the efficiencies averaged over certain chromosomes; therefore it may be useful to perform the entire detection process resulting in an observed abundance ratio for the mixed sample, do it also for the maternal sample, and then take the ratio of ratios. This ratio of ratios will be free of the chromosome bias; however, it will include errors in the measurements of the maternal sample.

[0203] Because the fraction of fetal cells can be small or even zero, the aneuploidy signal (the departure of the observed ratio from unity) may be weak even when fetal aneuploidy is present. An independent estimate of the fetal cell fraction, including a confidence estimate of whether measurable fetal DNA is present at all, is useful in interpreting the observed aneuploidy ratios. FIG. 7 illustrates allele signals reordered by rank. Assuming the mother has no more than two alleles at each locus, the magnitude of the third ranked allele is potentially a robust indicator of the presence of fetal DNA. Although measurement errors can artificially inflate the size of the third and fourth alleles, it is very unlikely to result in a bimodal distribution for the relative magnitude of the third allele with respect to the first two. Such a bimodal distribution is illustrated in FIG. 8. The secondary peak of this distribution occurs at a value approximately equal to the fraction of fetal cells. This is one way to determine the value of the variable f in the data model. The statistical confidence that the bimodality is real can be used to assign a confidence that fetal DNA was present in the mixed sample. Statistical tests for bimodality are discussed in M Y Cheng and P Hall, *J. R. Statist. Soc. B* (1998), 60 (Part 3) pp 579-589, and these authors prefer bootstrap based methods. Only if this confidence exceeds a threshold, say 99.9%, would an aneuploidy call be attempted. This threshold needs to be quite stringent to avoid the expensive mistake of declaring a fetus normal when in fact it is not. The estimated fetal cell fraction can be used to interpret the aneuploidy statistic: a large value of f and an observed aneuploidy ratio very close to unity would suggest no aneuploidy; a small value of f along with an aneuploidy ratio approximately equal to 1/f/2 would suggest trisomy, but it is still necessary to decide whether the observed aneuploidy is significantly different from unity and this requires an error model. A simple robust estimate of the error distribution could come from repeated processing of nominally identical samples.

Fitting of Data to the Model for Optimal Detection of Aneuploidy

[0204] The data model can be used to simultaneously recover estimates of the fraction of fetal cells, and efficient detection of aneuploidies in hypothesized chromosomes or chromosomal segments. This integrated approach should result in more reliable and sensitive declarations of aneuploidy.

[0205] Equations 1-3 actually represent five different models because of the ambiguity between $m_A$ and $m_B$ in the last of Equations 1 and 3. Testing for aneuploidy of Chromosomes 13, 18, and 21 would entail 5x5=25 different model variants that would be fit to the data.

[0206] The parameter values for the maternal allele identities are taken from the results for the maternal-only sample and the remaining parameters are fit to the data from the mixed sample. Because the number of parameters is very large even when the number of loci is large, a global optimization requires iterative search techniques. One possible approach is to do the following for each model variant:

1. Set f to 0 and solve for $A_f$ at each locus.

2. Set f to a value equal to the smallest fetal/maternal cell ratio for which fetal cells are likely to be detectable.

3. Solve for paternal allele(s) identities and strengths at each locus, one locus at a time, that minimize data-model residuals.

4. Fix the paternal alleles and adjust f to minimize residuals over all the data.

5. Now vary only the $A_f$ to minimize residuals. Repeat iv and v until convergence.
vi) Repeat iii through v until convergence.

The best overall fit of model to data is selected from among all the model variants. The best overall fit yields the values of \( f \) and \( \Lambda \), we will call \( f_{\text{max}}, \Lambda_{\text{max}} \). The likelihood of observing the data given \( f_{\text{max}} \) can be compared to the likelihood given \( f=0 \). The ratio is a measure of the amount of evidence for fetal DNA. A typical threshold for declaring fetal DNA would be a likelihood ratio of -1000 or more. The likelihood calculation can be approximated by a more familiar Chi-squared calculation involving the sum of squared residuals between the data and the model, where each residual is normalized by the expected rms error. This Chi-squared is a good approximation to the Log(likelihood) to the extent the expected errors in the data are Gaussian additive errors, or can be made so by some amplitude transformation of the data.

If based on the above determination of likelihood ratio it is decided that fetal DNA is not present, then the test is declared to be non-informative. If it is decided that fetal DNA is present, then the likelihoods of the data given the different data model types can be compared to declare aneuploidy. The likelihood ratios of aneuploid models (Equations 2 and 3) to the normal model (Equation 1) are calculated and these ratios are compared to a predefined threshold. Typically this threshold would be set so that in controlled tests all the trisomic cases would be declared aneuploid, and so that it would be expected that the vast majority (>99.9%) of all truly trisomic cases would be declared aneuploid by the test. Given a limited patient cohort size for test validation, one strategy to accomplish approximately the 99.9% detection rate is to increase the likelihood ratio threshold beyond that necessary to declare all the known trisomic cases in the validation set by a factor of 1000/N, where N is the number of trisomic cases in the validation set.

Error Modeling

The data contain small additive errors associated with the readout technology, multiplicative errors associated with DNA amplification and hybridization efficiencies being different from locus to locus and from allele to allele within a locus, and errors associated with imperfect specificity in the process. By including the many parameters \( \Lambda_k \) in the model, rather than a single scale parameter, the residuals will include allele-to-allele efficiency differences but not locus to locus differences. These tend to be multiplicative errors in the resulting observed allele strengths heights, i.e. two signals may be 20% different in strength although the starting concentrations of the alleles were identical. As a first approximation we can assume errors are random from allele to allele, and have relatively small additive errors, and larger Poisson and multiplicative error components. The magnitudes of these error components can be estimated from repeated processing of identical samples. The Chi-square residuals calculation for any data-model fit then can be supported with these modeled squared errors for any peak height or data bin. Alternatively, when using a single \( \Lambda \) parameter, the residuals will be larger and will contain a component which is correlated between alleles at the same locus. Calculation of likelihood will need to take this correlation into account.

---

**SEQUENCE LISTING**

<160> NUMBER OF SEQ ID NOS: 9

<210> SEQ ID NO 1
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 1
tcgagtgoat tecattcgc 19

<210> SEQ ID NO 2
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 2
atggaatggc atcaaacgga a

<210> SEQ ID NO 3
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:  

-continued

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic probe

<400> SEQUENCE: 3

tggtgtccatac 15
tccttt 15
tcatt 15

tggtgtccatac 15
tccttt 15
tcatt 15

<210> SEQ ID NO 4
<211> LENGTH: 65
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: Synthetic oligonucleotide

<400> SEQUENCE: 4

atgcagcaag gcacagacagtcacactagggaga gagaatggg 60
tcatt 65

<210> SEQ ID NO 5
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: Synthetic primer

<400> SEQUENCE: 5

atgcagcaag gcacagacagtcacactagggaga gagaatggg 60
tcatt 65

<210> SEQ ID NO 6
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: Synthetic primer

<400> SEQUENCE: 6

agaggggaga gagaatgggactcatt 23

<210> SEQ ID NO 7
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: Synthetic primer

<400> SEQUENCE: 7

caagagcaag actaagcaag gagaagc 25

<210> SEQ ID NO 8
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: Synthetic primer

<400> SEQUENCE: 8

ggcataaatttcactagggaga gagaatggga tcatt 35
1. A method for detecting fetal abnormality comprising:
   a. determining a ratio of abundance of maternal allele(s) to abundance of paternal allele(s) in genomic DNA from fetal cells enriched from a maternal blood sample using size-based separation.
2. The method of claim 1, wherein said genomic DNA comprises one or more SNPs.
3. The method of claim 2, wherein said SNP is informative.
4. (canceled)
5. The method of claim 2, wherein said one or more SNPs are in a single locus, different loci, single chromosome, or different chromosomes.
6. The method of claim 1, wherein said genomic DNA comprises a paternal allele that differs from both maternal alleles.
7. The method of claim 2, wherein a first region of genomic DNA is trisomic or suspected of being trisomic and a second region of genomic DNA is non-trisomic or suspected of not being trisomic.
8. The method of claim 7, further comprising the step of comparing said ratio of abundance of maternal allele(s) to abundance of paternal allele(s) in said first genomic DNA region with said ratio of abundance of maternal allele(s) to abundance of paternal allele(s) in said second genomic DNA region.
9. The method of claim 8, wherein an increase in paternal abundance in said first region is indicative of paternal trisomy.
10. The method of claim 8, wherein an increase in maternal abundance in said first region is indicative of maternal trisomy.
11. The method of claim 8, wherein said first genomic region and said second genomic region are on the same chromosome.
12. The method of claim 11, wherein an increase in paternal abundance or maternal abundance of one or more alleles is indicative of partial trisomy.
13-14. (canceled)
15. The method of claim 1, wherein said determining step comprises detecting an abundance of a nucleotide base at a SNP position.
16. The method of claim 2, wherein said SNP is detected using a DNA microarray, bead microarray, or high throughput sequencing.
17. The method of claim 1, further comprising enriching said sample for fetal cells prior to determining allele abundance.
18. (canceled)
19. A method for detecting fetal abnormality comprising:
   b. enriching a maternal blood sample for fetal cells using size-based separation;
   c. determining a ratio of allele abundance in said enriched maternal blood sample by comparing an abundance of one or more maternal alleles in a first genomic region with an abundance of one or more maternal alleles in a second genomic region,
   d. wherein said first genomic region is trisomic or suspected of trisomy, and
   e. wherein said second genomic region is non-trisomic.
20. The method of claim 19, wherein said maternal alleles in a first genomic region comprise an SNP.
21. The method of claim 19, wherein said maternal alleles in a second genomic region comprise an SNP.
22. The method of claim 19, further comprising comparing said ratio to a second ratio obtained for a control sample.
23. The method of claim 22, wherein said control sample comprises a diluted portion of said maternal blood sample.
24-25. (canceled)
26. The method of claim 19, further comprising ranking said alleles by allele abundance.
27. The method of claim 26, further comprising using said ranked alleles to determine an abundance of one or more paternal alleles.
28. The method of claim 27, further comprising comparing the abundance of said one or more paternal alleles to the abundance of said maternal alleles at one or more genetic regions.
29-31. (canceled)
32. The method of claim 19, wherein said first genomic region suspected of trisomy is selected from the group consisting of chromosomes 13, 18, 21 and X.
33. The method of claim 19, wherein said second genomic region that is non-trisomic is selected from the group consisting of chromosomes 13, 18, 21 and X.
34-35. (canceled)
36. A method for detecting fetal abnormality comprising:
   f. comparing an abundance of one or more maternal alleles in genomic DNA in a maternal blood sample with an abundance of one or more maternal alleles in said genomic DNA in a blood sample, wherein
   g. a first region of said genomic DNA is trisomic or suspected of trisomy, and,
   h. a second region of said genomic DNA is non-trisomic.
37-52. (canceled)