

the exception being the nesting protocol, and including the identity of the four samples. The ploidy calls for all 28 chromosomes in the set were called correctly with confidences above 99.7%. 7.6 million (97%) of reads mapped to the genome, and 6.3 million (80%) of the reads mapped to the targeted SNPs. The average depth of read was 751, and the median depth of
5 read was 396.

Experiment 11

In one experiment, three maternal plasma samples were split into five equal portions, and each portion was amplified using either 2,400 multiplexed primers (four portions) or 1,200
10 multiplexed primers (one portion) and amplified using a semi-nested protocol, for a total of 10,800 primers. After amplification, the portions were pooled together for sequencing. Details of Experiment 11 were very similar to Experiment 9, the exception being the nesting protocol, and the split and pool approach. The ploidy calls for all 21 chromosomes in the set were called correctly with confidences above 99.7%, except for one missed call where the confidence was
15 83%. 3.4 million reads mapped to targeted SNPs, the average depth of read was 404 and the median depth of read was 258.

Experiment 12

In one experiment, four maternal plasma samples were split into four equal portions,
20 and each portion was amplified using 2,400 multiplexed primers and amplified using a semi-nested protocol, for a total of 9,600 primers. After amplification, the portions were pooled together for sequencing. Details of Experiment 12 were very similar to Experiment 9, the exception being the nesting protocol, and the split and pool approach. The ploidy calls for all 28 chromosomes in the set were called correctly with confidences above 97%, except for one
25 missed call where the confidence was 78%. 4.5 million reads mapped to targeted SNPs, the average depth of read was 535 and the median depth of read was 412.

Experiment 13

In one experiment, four maternal plasma samples were prepared and amplified using a
30 9,600-plex triply hemi-nested protocol, for a total of 9,600 primers. Details of Experiment 12 were very similar to Experiment 9, the exception being the nesting protocol which involved three rounds of amplification; the three rounds involved 15, 10 and 15 STA cycles

respectively. The ploidy calls for 27 of 28 chromosomes in the set were called correctly with confidences above 99.9%, except for one that was called correctly with 94.6%, and one missed call with a confidence of 80.8%. 3.5 million reads mapped to targeted SNPs, the average depth of read was 414 and the median depth of read was 249.

5

Experiment 14

In one experiment 45 sets of cells were amplified using a 1,200-plex semi-nested protocol, sequenced, and ploidy determinations were made at three chromosomes. Note that this experiment is meant to simulate the conditions of performing pre-implantation genetic diagnosis on single-cell biopsies from day 3 embryos, or trophectoderm biopsies from day 5 embryos. 15 individual single cells and 30 sets of three cells were placed in 45 individual reaction tubes for a total of 45 reactions where each reaction contained cells from only one cell line, but the different reactions contained cells from different cell lines. The cells were prepared into 5 ul washing buffer and lysed the by adding 5 ul ARCTURUS PICOPURE lysis buffer (APPLIED BIOSYSTEMS) and incubating at 56°C for 20 min, 95°C for 10 min..

The DNA of the single/three cells was amplified with 25 cycles of STA (95°C for 10 min for initial polymerase activation, then 25 cycles of 95°C for 30s; 72°C for 10 s; 65°C for 1 min; 60°C for 8 min; 65°C for 3 min and 72°C for 30s; and a final extension at 72°C for 2 min) using 50 nM primer concentration of 1200 target-specific forward and tagged reverse primers.

The semi-nested PCR protocol involved three parallel second amplification of a dilution of the first STAs product for 20 cycles of STA (95°C for 10 min for initial polymerase activation, then 15 cycles of 95°C for 30s; 65°C for 1 min; 60°C for 5 min; 65°C for 5 min and 72°C for 30s; and a final extension at 72°C for 2 min) using reverse tag specific primer concentration of 1000 nM, and a concentration of 60 nM for each of 400 target-specific nested forward primers. In the three parallel 400-plex reactions the total of 1200 targets amplified in the first STA were thus amplified.

An aliquot of the STA products was then amplified by standard PCR for 15 cycles with 1 uM of tag-specific forward and barcoded reverse primers to generate barcoded sequencing libraries. An aliquot of each library was mixed with libraries of different barcodes and purified using a spin column.

In this way, 1,200 primers were used in the single cell reactions; the primers were designed to target SNPs found on chromosomes 1, 21 and X. The amplicons were then sequenced using an ILLUMINA GAIIX sequencer. Per sample, approximately 3.9 million reads were generated by the sequencer, with 500,000 to 800,000 million reads mapping to the genome (74% to 94% of all reads per sample).

Relevant maternal and paternal genomic DNA samples from cell lines were analyzed using the same semi-nested 1200-plex assay pool with a similar protocol with fewer cycles and 1200-plex second STA, and sequenced.

The sequencing data was analyzed using informatics methods disclosed herein and the ploidy state was called at the three chromosomes for the samples.

Figure 24 shows normalized depth of read ratios (vertical axis) for six samples at three chromosomes (1 = chrom 1; 2 = chrom 21; 3 = chrom X). The ratios were set to be equal to the number of reads mapping to that chromosome, normalized, and divided by the number of reads mapping to that chromosome averaged over three wells each comprising three 46XY cells. The three sets of data points corresponding to the 46XY reactions are expected to have ratios of 1:1. The three sets of data points corresponding to the 47XX+21 cells are expected to have ratios of 1:1 for chromosome 1, 1.5:1 for chromosome 21, and 2:1 for chromosome X.

Figure 25 shows allele ratios plotted for three chromosomes (1, 21, X) for three reaction. The reaction in the lower left shows a reaction on three 46XY cells. The left region are the allele ratios for chromosome 1, the middle region are the allele ratios for chromosome 21, and the right region are the allele ratios for chromosome X. For the 46XY cells, for chromosome 1 we expect to see ratios of 1, 0.5 and 0, corresponding to AA, AB and BB SNP genotypes. For the 46XY cells, for chromosome 21 we expect to see ratios of 1, 0.5 and 0, corresponding to AA, AB and BB SNP genotypes. For the 46XY cells, for chromosome X we expect to see ratios of 1 and 0, corresponding to A, and B SNP genotypes. The reaction in the lower right shows a reaction on three 47XX+21 cells. The allele ratios are segregated by chromosome as in the lower left graph. For the 47XX+21 cells, for chromosome 1 we expect to see ratios of 1, 0.5 and 0, corresponding to AA, AB and BB SNP genotypes. For the 47XX+21 cells, for chromosome 21 we expect to see ratios of 1, 0.67, 0.33 and 0, corresponding to AAA, AAB, ABB and BBB SNP genotypes. For the 47XX+21 cells, for chromosome X we expect to see ratios of 1, 0.5 and 0, corresponding to AA, AB, and BB SNP genotypes. The plot in the upper right was made on a reaction comprising 1 ng of genomic

DNA from the 47XX+21 cell line. **Figure 26** shows the same graphs as in Figure 25, but for reactions performed on only one cell. The left graph was a reaction that contained a 47XX+21 cell, and the right graph was for a reaction that contained a 46XX cell.

From the graphs shown in **Figure 25** and **Figure 26**, it is visually apparent that there are two clusters of dots for chromosomes where we expect to see ratios of 1 and 0; three clusters of dots for chromosomes where we expect to see ratios of 1, 0.5, and 0, and four clusters of dots for chromosomes where we expect to see ratios of 1, 0.67, 0.33 and 0. The parental support algorithm was able to make correct calls on all of the three chromosomes for all of the 45 reactions.

Experiment 15

In one experiment, maternal plasma samples were prepared and amplified using a hemi-nested 19,488-plex protocol. The samples were prepared in the following way: up to 20 mL of maternal blood were centrifuged to isolate the buffy coat and the plasma. The genomic DNA in the maternal sample was prepared from the buffy coat and paternal DNA was prepared from a blood sample or saliva sample. Cell-free DNA in the maternal plasma was isolated using the QIAGEN CIRCULATING NUCLEIC ACID kit and eluted in 50 uL TE buffer according to manufacturer's instructions. Universal ligation adapters were appended to the end of each molecule of 40 uL of purified plasma DNA and libraries were amplified for 9 cycles using adaptor specific primers. Libraries were purified with AGENCOURT AMPURE beads and eluted in 50 ul DNA suspension buffer.

6 ul of the DNA was amplified with 15 cycles of STAR 1 (95°C for 10 min for initial polymerase activation, then 15 cycles of 96°C for 30s; 65°C for 1 min; 58°C for 6 min; 60°C for 8 min; 65°C for 4 min and 72°C for 30s; and a final extension at 72°C for 2 min) using 7.5 nM primer concentration of 19,488 target-specific tagged reverse primers and one library adaptor specific forward primer at 500 nM.

The hemi-nested PCR protocol involved a second amplification of a dilution of the STAR 1 product for 15 cycles (STAR 2) (95°C for 10 min for initial polymerase activation, then 15 cycles of 95°C for 30s; 65°C for 1 min; 60°C for 5 min; 65°C for 5 min and 72°C for 30s; and a final extension at 72°C for 2 min) using reverse tag concentration of 1000 nM, and a concentration of 20 nM for each of 19,488 target-specific forward primers.

An aliquot of the STAR 2 products was then amplified by standard PCR for 12 cycles with 1 uM of tag-specific forward and barcoded reverse primers to generate barcoded sequencing libraries. An aliquot of each library was mixed with libraries of different barcodes and purified using a spin column.

5 In this way, 19,488 primers were used in the single-well reactions; the primers were designed to target SNPs found on chromosomes 1, 2, 13, 18, 21, X and Y. The amplicons were then sequenced using an ILLUMINA GAIIX sequencer. For plasma samples, approximately 10 million reads were generated by the sequencer, with 9.4-9.6 million reads mapping to the genome (94-96 %), and of those, 99.95 % mapped to targeted SNPs with a mean depth of read
10 of 460 and a median depth of read of 350. For comparison, a perfectly even distribution would be: 10M reads / 19,488 targets = 513 reads/target. For primer-dimers, 30,000 reads were from sequenced primer-dimers (0.3% of the reads generated by the sequencer). For genomic samples, 99.4-99.7% of the reads mapped to the genome, of those, 99.99% of the mapped to targeted SNPs, and 0.1% of the reads generated by the sequencer were primer-dimers.

15 For plasma samples with 10 million sequencing reads, typically at least 19,350 of the 19,488 targeted SNPs (99.3 %) are amplified and sequenced. For DNA samples with 2M sequencing reads, typically at least 19,000 targeted SNPs (97.5%) are amplified and sequenced. The lower number may be due to sampling noise since the number of reads is lower and the sequencer misses some of the amplified products. If desired, the number of
20 sequencing reads can be increased to increase the number of targeted SNPs that are amplified and sequenced.

Relevant maternal and paternal genomic DNA samples amplified using a semi-nested 19,488 outer forward primers and tagged reverse primers at 7.5 nM in the STAR 1. Thermocycling conditions and composition of STAR 2, and the barcoding PCR were the same
25 as for the hemi-nested protocol.

The average fetal fraction for 407 samples was found to be 14.8%. The sequencing data was analyzed using informatics methods disclosed herein and the ploidy state was called at four chromosomes (13, 18, 21, Y) for the fetuses whose DNA was present in 378 of the 407 maternal plasma samples, and at chromosome X in 375 of the 407 maternal plasma samples.
30 The ploidy calls for all 1,887 chromosomes in the set were called correctly with confidences above 90%. 1882 of the 1887 calls were above 95%; and 1,862 of the 1,887 calls were called with confidences above 99%.

A similar control experiment was performed using water instead of DNA extracted from plasma in the plasma PCR protocol. Based on six such trials of an experiment, 5-6% of the sequenced reads were primer-dimers. Other sequenced reads were due to background noise. This experiment demonstrates that even in the absence of a nucleic acid sample with
 5 target loci for the primers to hybridize to (rather than hybridizing to other primers and forming amplified primer dimers) few primer dimers are formed.

Experiment 16

The following experiment illustrates an exemplary method for designing and selecting
 10 a library of primers that can be used in any of the multiplexed PCR methods of the invention. The goal is to select primers from an initial library of candidate primers that can be used to simultaneously amplify a large number of target loci (or a subset of target loci) in a single reaction. For an initial set of candidate target loci, primers did not have to be designed or selected for each target locus. Preferably, primers are designed and selected for a large portion
 15 of the most desirable target loci.

Step 1

A set of candidate target loci (such as SNPs) were selected based on publically available information about desired parameters for the target loci, such as frequency of the
 20 SNPs within a target population or heterozygosity rate of the SNPs (worldwide web at ncbi.nlm.nih.gov/projects/SNP/; Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001 Jan 1;29(1):308-11, which are each incorporated by reference in its entirety). For each candidate locus, one or more PCR primer pairs were designed using the Primer3 program (the worldwide web at
 25 primer3.sourceforge.net; libprimer3 release 2.2.3, which is hereby incorporated by reference in its entirety). If there were no feasible designs for PCR primers for a particular target locus, then that target locus was eliminated from further consideration.

If desired, a “target locus score” (higher score representing higher desirability) can be calculated for most or all of the target loci, such as a target locus score calculated based on a
 30 weighted average of various desired parameters for the target loci. The parameters may be assigned different weights based on their importance for the particular application that the primers will be used for. Exemplary parameters include the heterozygosity rate of the target

locus, the disease prevalence associated with a sequence (e.g., a polymorphism) at the target locus, the disease penetrance associated with a sequence (e.g., a polymorphism) at the target locus, the specificity of the candidate primer(s) used to amplify the target locus, the size of the candidate primer(s) used to amplify the target locus, and the size of the target amplicon.

5

Step 2

A thermodynamic interaction score was calculated between each primer and all primers for all other target loci from Step 1 (see, e.g., Allawi, H. T. & SantaLucia, J., Jr. (1998), "Thermodynamics of Internal C-T Mismatches in DNA", *Nucleic Acids Res.* 26, 2694-2701; 10 Peyret, N., Seneviratne, P. A., Allawi, H. T. & SantaLucia, J., Jr. (1999), "Nearest-Neighbor Thermodynamics and NMR of DNA Sequences with Internal A-A, C-C, G-G, and T-T Mismatches", *Biochemistry* 38, 3468-3477; Allawi, H. T. & SantaLucia, J., Jr. (1998), "Nearest-Neighbor Thermodynamics of Internal A-C Mismatches in DNA: Sequence Dependence and pH Effects", *Biochemistry* 37, 9435-9444.; Allawi, H. T. & SantaLucia, J., Jr. 15 (1998), "Nearest Neighbor Thermodynamic Parameters for Internal G-A Mismatches in DNA", *Biochemistry* 37, 2170-2179; and Allawi, H. T. & SantaLucia, J., Jr. (1997), "Thermodynamics and NMR of Internal G-T Mismatches in DNA", *Biochemistry* 36, 10581-10594; MultiPLX 2.1 (Kaplinski L, Andreson R, Puurand T, Remm M. MultiPLX: automatic grouping and evaluation of PCR primers. *Bioinformatics*. 2005 Apr 15;21(8):1701-2, which 20 are each hereby incorporated by reference in its entirety). This step resulted in a 2D matrix of interaction scores. The interaction score predicted the likelihood of primer-dimers involving the two interacting primers. The score was calculated as follows:

$$\text{interaction_score} = \max(-\text{deltaG_2}, 0.8 * (-\text{deltaG_1}))$$

25

where

deltaG_2 = Gibbs energy (energy required to break the dimer) for a dimer that is extensible by PCR on both ends, i.e., the 3' end of each primer anneals to the other primer; and

deltaG_1 = Gibbs energy for a dimer that is extensible by PCR on at least one end.

30

Step 3:

For each target locus, if there was more than one primer-pair design, then one design was selected using the following method:

- 1 For each primer-pair design for the locus, find the worst-case (highest) interaction score for the two primers in that design and all primers from all designs for all other target loci.
- 2 Pick the design with the best (lowest) worst-case interaction score.

Step 4

A graph was built such that each node represented one locus and its associated primer-pair design (e.g., a Maximal Clique problem). One edge was created between every pair of nodes. A weight was assigned to each edge equal to the worst-case (highest) interaction score between the primers associated with the two nodes connected by the edge.

Step 5

If desired, for every pair of designs for two different target loci where one of the primers from one design and one of the primers from the other design would anneal to overlapping target regions, an additional edge was added between the nodes for the two design. The weight of these edges was set equal to the highest weight assigned in Step 4. Thus, Step 5 prevents the library from having primers that would anneal to overlapping target regions, and thus interfere with each other during a multiplex PCR reaction.

Step 6

An initial interaction score threshold was calculated as follows:

$$\text{weight_threshold} = \max(\text{edge_weight}) - 0.05 * (\max(\text{edge_weight}) - \min(\text{edge_weight}))$$

where

$\max(\text{edge_weight})$ is the maximum edge weight in the graph; and

$\min(\text{edge_weight})$ is the minimum edge weight in the graph.

The initial bounds for the threshold were set as follows:

$$\text{max_weight_threshold} = \max(\text{edge_weight})$$

$$\text{min_weight_threshold} = \min(\text{edge_weight})$$

Step 7

A new graph was constructed consisting of the same set of nodes as the graph from Step 5, only including edges with weights that exceed `weight_threshold`. Thus, step ignores interactions with scores equal to or below `weight_threshold`.

5 *Step 8*

Nodes (and all of the edges connected to the removed nodes) were removed from the graph of Step 7 until there were no edges left. Nodes were removed by applying the following procedure repeatedly:

- 1 Find the node with the highest degree (highest number of edges). If there is more than
10 one then pick one arbitrarily.
- 2 Define the set of nodes consisting of the node picked above and all of the nodes
connected to it, but excluding any nodes that have degree less than the node picked
above.
- 3 Choose the node from the set that has the lowest target locus score (lower score
15 representing lower desirability) from Step 1. Remove that node from the graph.

Step 9

If the number of nodes remaining in the graph satisfies the required number of target loci for the multiplexed PCR pool (within an acceptable tolerance), then the method was
20 continued at Step 10.

If there were too many or too few nodes remaining in the graph, then a binary search was performed to determine what threshold values would result in the desired number of nodes remaining in the graphs. If there were too many nodes in the graph then, the weight threshold bounds were adjusted as follows:

25 `max_weight_threshold = weight_threshold`

Otherwise (if there are too few nodes in the graph), then the weight threshold bounds were adjusted as follows:

`min_weight_threshold = weight_threshold`

Then, the weight threshold was adjusted follows:

30 `weight_threshold = (max_weight_threshold + min_weight_threshold) / 2`

Steps 7-9 were repeated.

Step 10

The primer-pair designs associated with the nodes remaining in the graph were selected for the library of primers. This primer library can be used in any of the methods of the invention.

5 If desired, this method of designing and selecting primers can be performed for primer libraries in which only one primer (instead of a primer pair) is used for amplification of a target locus. In this case, a node presents one primer per target locus (rather than a primer pair).

10 *Experiment 17*

Figure 27 is a graph comparing two primer libraries designed using the methods of the invention. This graph shows the number of loci with a particular minor allele frequency that are targeted by each primer library. During the selection of the “new pool” library, more primers were retained. This library enables the amplification of more target loci, especially target loci with relatively large minor allele frequencies (which are the more informative alleles for some method of the invention, such as for detecting fetal chromosomal abnormalities).

These primer libraries were used in the following multiplex PCR method. Blood (20-40 mL) was collected from each subject into two to four CELL-FREE™ DNA tubes (Streck). Plasma (a minimum of 7 mL) was isolated from each sample via a double centrifugation protocol of 2,000 g for 20 min, followed by 3,220 g for 30 min, with supernatant transfer following the first spin. cfDNA was isolated from 7-20 mL plasma using the QIAGEN QIAamp Circulating Nucleic Acid kit and eluted in 45 uL TE buffer. Pure maternal genomic DNA was isolated from the buffy coat obtained following the first centrifugation, and pure paternal genomic DNA was prepared similarly from a blood, saliva or buccal sample.

Maternal cfDNA, maternal genomic DNA, and paternal genomic DNA samples were pre-amplified for 15 cycles using 11,000 target-specific assays and an aliquot was transferred to a second PCR reaction of 15 cycles using nested primers. Finally, samples were prepared for sequencing by adding barcoded tags in a third 12-cycle round of PCR. Thus, 11,000 targets were amplified in a single reaction; the targets included SNPs found on chromosomes 13, 18, 21, X, and Y. The amplicons were then sequenced using an ILLUMINA GAIIx or HISEQ

sequencer. Parental genotypes were sequenced at a lower read depth (~20% of cfDNA read depth) than the fetal genotypes.

Experiment 18

5 If desired, the size and quantity of the PCR products can be analyzed using standard methods, such as the use of the Agilent Technologies 2100 Bioanalyzer (**Figure 28A-M**). For example, direct PCR methods described herein without nesting were used in 2,400-plex (**Figures 28B-28G**) and 19,488-plex experiments (**Figures 28H to 28M**). The amount of primer was 10 nM for **Figures 28B-28D and 28H to 28J**. The amount of primer was 1 nM for
10 **Figures 28E-28G and 28K to 28M**. The amount of input DNA was 24 ng for **Figures 28B, 28E, 28H, and 28K**; 80 ng for **Figures 28C, 28F, 28I, and 28L**; and 250 ng for **Figures 28D, 28G, 28J, and 28M**. More input DNA resulted in a greater proportion of the desired 180 base pair product. The peak at 140 base pairs is a primer dimer product.

15 *Experiment 19*

A proof-of-principle study demonstrated the detection of T13, T18, T21, 45,X, and 47,XXY with equally high accuracies across all chromosomes.

Patients

20 Pregnant couples were enrolled at specific prenatal care centers under protocols approved by an Institutional Review Board pursuant to local laws. Inclusion criteria were at least 18 years of age, gestational age of at least nine weeks, singleton pregnancies, and signed informed consent. Blood samples were drawn from pregnant mothers, and a blood or buccal sample was collected from the father. Samples from 2 pregnancies with T13 (Patau
25 Syndrome), 2 with T18 (Edwards Syndrome), 2 with T21 (Down's Syndrome), 2 with 45,X, 2 with 47,XXY, and 90 normal pregnancies were selected prior to testing from a cohort of ~500 women to test which chromosomal abnormalities the method detects. Normal fetal karyotype was confirmed by molecular karyotyping for the samples where post-birth child tissue was available. Euploid sample were drawn prior to invasive testing from low-risk women.
30 Aneuploid samples were drawn at least 7 days after invasive testing and aneuploidy was confirmed via cytogenetic karyotyping or fluorescence *in situ* hybridization at independent laboratories.

Sample Preparation and Multiplex PCR

For the data in **Figures 30A-E, 30G, 30H, and 31A-31G**, sample preparation and 19,488-plex-PCR were performed as described in Experiment 15. For the data in **Figure 30F**,
5 sample preparation and 11,000-plex-PCR were performed as described in Experiment 17.

Methodology and Data Analysis

The algorithm considers parental genotypes and crossover frequency data (such as data from the HapMap database) to calculate expected allele distributions for 19,488 polymorphic
10 loci for a very large number possible fetal ploidy states, and at various fetal cfDNA fractions. (**Figures 29A-29C**). Unlike allele ratio based-methods, it also takes into account linkage disequilibrium, and uses non-Gaussian data models to describe the expected distribution of allele measurements at a SNP given observed platform characteristics and amplification biases. It then compares the various predicted allele distributions to the actual allelic distributions as
15 measured in the cfDNA sample (**Figure 29C**), and calculates the likelihood of each hypothesis (monosomy, disomy, or trisomy, for which there are numerous hypotheses based on the various potential crossovers) based on the sequencing data. The algorithm sums the likelihoods of each individual monosomy, disomy, or trisomy hypotheses (**Figure 29D**), and calls the ploidy state with the maximum overall likelihood as the copy number and fetal fraction
20 (**Figure 29E**). Although laboratory researchers were not blinded to sample karyotype, the algorithm called the ploidy states without human intervention and was blind to the truth.

Data Interpretation

Graphical Representations of the Generated Data

25 To determine the ploidy state of a chromosome of interest, the algorithm considers the distribution of sequence counts from each of two possible alleles at 3,000 to 4,000 SNPs per chromosome. It is important to note that the algorithm makes ploidy calls using an approach that does not lend itself to visualization. Thus, for the purposes of illustration, the data is displayed here in a simplified fashion as ratios of the two most likely alleles, labeled as A and
30 B, so that the relevant trends can be more readily visualized. This simplified illustration does not take into account some of the features of the algorithm. For example, two important aspects of the algorithm that are not possible to illustrate with a method of visualization that

displays allele ratios are: 1) the ability to leverage linkage disequilibrium, *i.e.* the influence that a measurement at one SNP has on the likely identity of a neighboring SNP, and 2) the use of non-Gaussian data models that describe the expected distribution of allele measurements at a SNP given platform characteristics and amplification biases. Also note that the algorithm only
5 considers the two most common alleles at each SNP, ignoring other possible alleles.

The graphical representations in **Figure 30A-30H** include samples for which two, one, or three fetal chromosomes are present. Generally, these indicate euploidy (**Figures 30A-30C**) monosomy (**Figure 30D**), and trisomy (**Figures 30E-30H**), respectively. In all plots, each spot represents a single SNP, where the targeted SNPs are plotted sequentially from left to
10 right for one chromosome along the horizontal axes. The vertical axes indicate the number of reads for the A allele as a fraction of the total number of reads for both the A and B alleles for that SNP. Note that the measurements are made on total cfDNA isolated from maternal blood, and the cfDNA includes both maternal and fetal cfDNA; thus, each spot represents the combination of the fetal and maternal DNA contribution for that SNP. Therefore, increasing
15 the proportion of maternal cfDNA from 0% to 100% will gradually shift some spots up or down within the plots, depending on the maternal and fetal genotype. This is described in more detail below with the corresponding plots.

The spots are color-coded according to maternal genotype, as maternal genotype contributes more to the localization of each spot and the majority of trisomies are maternally-
20 inherited; this assists in visualizing ploidy states. Specifically, SNPs for which the maternal genotype is AA are indicated in red, those for which the maternal genotype is AB are indicated in green, and those for which the maternal genotype is BB are indicated in blue.

In all cases, SNPs that are homozygous for the A allele (AA) in both the mother and the fetus are found tightly associated with the upper limit of the plots, as the fraction of A allele
25 reads is high because there should be no B alleles present. Conversely, SNPs that are homozygous for the B allele in both the mother and the fetus are found tightly associated with the lower limit of the plots, as the fraction of A allele reads is low because there should be only B alleles. The spots that are not tightly associated with the upper and lower limits of the plots represent SNPs for which the mother, the fetus, or both are heterozygous; these spots are
30 useful for identifying fetal ploidy, but can also be informative for determining paternal versus maternal inheritance. These spots segregate based on both maternal and fetal genotypes and fetal fraction, and as such the precise position of each individual spot along the y-axis depends

on both stoichiometry and fetal fraction. For example, loci where the mother is AA and the fetus is AB are expected to have a different fraction of A allele reads, and thus different positioning along the y-axis, depending on the fetal fraction.

5 *Two Chromosomes Present*

Figures 30A-30C depict data that indicate the presence of two chromosomes when the sample is entirely maternal (no fetal cfDNA present, **Figure 30A**), contains a moderate fetal cfDNA fraction (**Figure 30B**), or contains a high fetal cfDNA fraction (**Figure 30C**).

Figure 30A shows data obtained from cfDNA isolated from the blood of a non-
 10 pregnant woman. When there is no fetal cfDNA present and the sample contains only maternal cfDNA, the plots represent purely the euploid maternal genotype; the hallmark pattern includes “clusters” of spots: a red cluster tightly associated with the top of the plot (SNPs where the maternal genotype is AA), a blue cluster tightly associated with the bottom of the plot (SNPs where the maternal genotype is BB), and a single, centered green cluster (SNPs where the
 15 maternal genotype is AB).

When fetal cfDNA is present, the location of the spots shifts such that the clusters segregate into discrete “bands”. Note that for samples with a fetal fraction of 0%, the groupings of spots are referred to as “clusters” (as in **Figure 30A**), and for all samples with a fetal fraction of >0%, the groupings of spots are referred to as “bands” (as in **Figures 30B-30J**). If the fetal fraction is high enough, these discrete bands will be readily visible. Specifically, **Figures 30B and 30C** demonstrate the characteristic pattern associated with two fetal chromosomes present at moderate and high fetal fractions, respectively. This pattern includes three central green bands that correspond to SNPs that are heterozygous in the mother, and two “peripheral” bands each at both the top (red) and bottom (blue) of the plots
 20 that correspond to SNPs that are homozygous in the mother.
 25

Figure 30B shows data obtained from cfDNA isolated from a plasma sample from a woman carrying a euploid fetus and with a 12% fetal cfDNA fraction. Here, the clusters of spots tightly associated with the top and bottom of the plot segregate into two discrete bands each: one red and one blue external peripheral band that remains tightly associated with the
 30 upper or lower limit of the plots, and one red and one blue internal peripheral band that has separated from the limits of the plots. These internal peripheral bands, centered around 0.92 and 0.08, represent SNPs for which the maternal genotype is AA and the fetal genotype is AB

(indicated in red), and SNPs for which the maternal genotype is BB and the fetal genotype is AB (indicated in blue), respectively. The center cluster of green spots broadens, but at this fetal fraction the segregation into distinct bands is not readily visible.

At a high fetal cfDNA fraction, the typical pattern that indicates the presence of two chromosomes (a trio of green bands as well as two red and two blue peripheral bands) is readily apparent. **Figure 30C** displays data obtained from a plasma sample from a woman carrying a euploid fetus at a fetal cfDNA fraction of 26%. Here, the peripheral bands have separated such that the internal band has shifted towards the center of the plot due to the altered levels of B alleles from the increased fetal cfDNA fraction. Significantly, at higher fetal fractions, the separation of the center green cluster into three distinct bands is now readily apparent. This central trio of bands, in this case clustering around 0.37, 0.50 and 0.63, corresponds to those SNPs where the maternal genotype is AB, and the fetal genotype is AA (top), AB (middle) and BB (bottom).

These hallmark patterns, namely three green bands and four peripheral bands (two red and two blue), indicate the presence of two chromosomes, as in autosomal euploidy or for the X chromosome in a female (XX) fetus.

One Chromosome Present

When the fetus only inherits a single chromosome, and thus only inherits a single allele, heterozygosity of the fetus is not possible. As such, the only possible fetal SNP identities are A or B. Thus, maternally-inherited monosomic chromosomes have a characteristic pattern of two central green bands that represent SNPs for which the mother is heterozygous, and only have single peripheral red and blue bands that represent SNPs for which the mother is homozygous, and which remain tightly associated with the upper and lower limits of the plots (1 and 0), respectively (**Figure 30D**). Note the absence of internal peripheral bands. This pattern indicates the presence of one chromosome, as in maternally-inherited autosomal monosomy, or for the X chromosome in a male (XY) fetus.

Three Chromosomes Present

Trisomic chromosomes have three characteristic patterns. The first pattern indicates maternally-inherited meiotic trisomy, a meiotic error where the fetus inherited two homologous, non-identical chromosomes from the mother (**Figure 30E**); this pattern includes

two central green bands with two each of the peripheral red and blue bands. The second pattern indicates paternally-inherited meiotic trisomy, where the fetus inherited two homologous, non-identical chromosomes from the father (**Figure 30F**); this pattern includes four central green bands and three each of the peripheral red and blue bands. The third pattern indicates either

5 maternally- (**Figure 30G**) or paternally-inherited (**Figure 30H**) mitotic trisomy, a mitotic error where the fetus inherited two identical chromosomes from either the mother or the father; this pattern includes four central green bands with two each of the peripheral red and blue bands. Maternally- and paternally-inherited mitotic trisomies can be distinguished by the placement of

10 the flanking red and blue bands, such that the red and blue internal peripheral bands (those not associated with the limits of the plots) are closer to the center in paternally-inherited mitotic trisomy. This is due to the paternal contribution of identical chromosomes. Note that our previous results indicate that at the blastomere stage, 66.7% of maternally-inherited trisomies are meiotic, and that only 10.2% of trisomies are paternally-inherited.

For the Y chromosome, the PS method considers a different set of hypotheses: zero,

15 one, or two chromosomes present. As there is no maternal contribution to the sequence reads at each locus and because heterozygous loci are not possible (cases of two Y chromosomes necessarily involve two identical chromosomes), the bands remain tightly associated with the top (A alleles) or the bottom (B alleles) of the plot (data not shown), and analysis is greatly simplified, relying on quantitative allele count data. Note that since the method interrogates

20 SNPs, it uses homologous non-recombinant SNPs from the Y chromosome, thus obtaining data on both X and Y for one probe pair.

Identifying Aneuploidies

Identification of autosomal aneuploidies using this plot-based visualization method is

25 straightforward given a sufficient fetal fraction, and requires only identifying plots for which there are an abnormal number of chromosomes present, as described above. Combining the knowledge of copy number of the X and Y chromosomes identifies whether sex chromosome aneuploidies are present. Specifically, plots representing a fetus with a 47,XXX genotype will have a typical “three-chromosome” pattern, and plots representing a fetus with a 47,XXY

30 genotype will have the typical “two-chromosome” pattern for the X chromosome, but will also have allele reads indicating the presence of one Y chromosome. The method is similarly able to call 47,XYY, where a “one chromosome” pattern indicates the presence of a single X

chromosome, and allele reads indicate the presence of two Y chromosomes. A fetus with a 45,X genotype will have the typical “one-chromosome” pattern for the X chromosome, and data indicating zero Y chromosomes.

5 *Effects of Fetal Fraction*

As discussed above, the number of sequence reads from the fetus contributes to the precise location of each spot along the y-axis in the plots. As fetal fraction will affect the proportion of reads that originate from the fetus and the mother, it will also affect the positioning of each spot. At a high fraction of fetal cfDNA (generally above ~20%), as in
 10 **Figures 30C-30E and Figures 30G and 30H**, it is readily apparent that although the spots cluster based mainly on maternal genotype, the presence of fetal DNA from alleles whose genotype is distinct from the maternal genotype shift the clusters into multiple, distinct bands. However, as the fetal fraction decreases (as in **Figures 30B and 30F**), the spots regress towards the poles and center of the plot, resulting in tighter clusters. Specifically, the set of
 15 peripheral red bands, where the maternal genotype is AA, regress towards the top of the plot; the set of peripheral blue bands, where the maternal genotype is BB, regress towards the bottom; the set of central green bands, where the mother is heterozygous, condense into a single cluster at the center of the plot (compare **Figures 30B and 30C**). Although aneuploidy is not readily apparent by eye using this visualization technique for low fetal fraction cases, the
 20 algorithm is able to identify ploidy states with a very low fetal fraction, such as 3% fetal fraction. It is able to do this because the statistical technique compares the observed data to very precise data models that predict the allele distributions for a given sample parameter set (including copy number, parental genotypes, and fetal fraction, for example). Data model precision is critical in low fetal fraction cases, as the differences between the allele
 25 distributions for different ploidy states are proportional to the fetal fraction. In addition, the algorithm is able to determine when a data set does not contain enough data to make a confident fetal ploidy determination.

Results

30 Sequencing reads that mapped to targeted SNPs were deemed to be informative and were used by the algorithm. More than 95% of targeted loci were observed in the sequencing results. The plots for visualizing key ploidy calls are depicted in **Figure 31A-31G. Figure**

31A indicates a euploid sample. Here, chromosomes 13, 18, and 21 have the typical “two chromosome” pattern (as described herein). This includes a trio of center green bands, and two red and two blue peripheral bands. This, together with the two center green bands for the X chromosome and the presence of Y chromosome bands along the plots’ peripheries, indicate a euploid XY genotype.

The most prevalent autosomal trisomies, T13, T18, and T21, are indicated by the plots in **Figures 31B, 31C, and 31D**, respectively. Specifically, **Figure 31B** depicts a T13 sample. Here, chromosomes 18 and 21 display the typical “two chromosome” pattern, chromosome X displays the typical “one chromosome” pattern, and there are reads from the Y chromosome. Together, this indicates disomy at chromosomes 18 and 21, and identifies a fetal XY genotype. However, chromosome 13 depicts a typical “three chromosome” pattern – specifically. Similarly, **Figure 31C** depicts a T18 sample, and **Figure 31D** depicts a T21 sample.

The method is also able to detect sex chromosome aneuploidies, including 45,X (**Figure 31E**), 47,XXY (**Figure 31F**), and 47,XYY (**Figure 31G**). Note that the method is calling copy number at chromosomes 13, 18, 21, X, and Y; the overall chromosome number is reported assuming disomy at the remaining chromosomes. The X chromosome regions of the plot depicting a 45,X sample reveals the presence of a single chromosome. However, the lack of reads from the Y chromosome, coupled with the “two chromosome” pattern for chromosomes 13, 18, and 21, indicate a 45,X genotype. Conversely, the 47,XXY samples generate a plot revealing the presence of two X chromosomes. The data also revealed reads for alleles from the Y chromosome. Together with the presence of two copies of chromosomes 13, 18, and 21, this indicates a 47,XXY genotype. A 47,XYY genotype is indicated by the presence of a “one chromosome” pattern for the X chromosome, and reads indicating the presence of two Y chromosomes.

Discussion

This method detected T13, T18, T21, 45,X, 47,XXY, and 47,XYY non-invasively from maternal blood. This method interrogates cfDNA from maternal plasma by targeted multiplex PCR amplification and high-throughput sequencing of 19,488 SNPs. This, coupled with the method’s sophisticated informatics analyses that take into account parental genotypic information and numerous sample parameters, including fetal fraction and DNA quality, more robustly detects the fetal signal and makes highly accurate ploidy calls at all of the five

chromosomes implicated in the seven most common types of at-birth aneuploidy (T13, T18, T21, 45,X, 47,XXX, 47,XXY, and 47,XYY). This method offers a number of clinical advantages over previous methods, including and most significantly greater clinical coverage and sample-specific calculated accuracies (analogous to a personalized risk score).

5

Increased Clinical Coverage

This method offers approximately a two-fold increase in aneuploidy coverage compared to clinically available NIPT methodologies, given its ability to accurately detect autosomal trisomies and sex chromosome aneuploidies. The method presented here is the only noninvasive test that calls ploidy at the sex chromosomes with high accuracy. Prior DNA mixing experiments and separate plasma samples analyzed in our experimental assays suggest that this method will detect a larger cohort of sex chromosome anomalies, including 47,XXX. The method presented here also detects aneuploidies at chromosomes 13, 18, and 21 with high sensitivities and specificities, and with appropriate primer design is expected to be able to detect copy number at the remaining chromosomes as well.

15

Sample-Specific Calculated Accuracies

Significantly, this method calculates a sample-specific accuracy for ploidy calls on each chromosome in each sample. Accuracies calculated by this method are expected to significantly lower the rate of incorrect calls by identifying and flagging individual samples that have poor quality DNA or low fetal fractions that are likely to result in a poor accuracy test result. By contrast, massively parallel shotgun sequencing (MPSS)-based methods produce a positive or negative call using a single-hypothesis rejection test, and their accuracy estimate is based on a published study cohort rather than on the characteristics of the individual sample, which are assumed to have the same accuracy as the cohort. However, individual accuracies for samples with parameters in the tail of the cohort distribution may differ significantly. This is exacerbated at low fetal fractions, as in early gestational age, or for samples with low DNA quality. These samples are generally not identified and flagged for follow-up, which can result in missed calls. The present method, however, takes into account many parameters, including fetal fraction and a number of DNA quality metrics, to make each chromosome copy number call, calculating a sample-specific accuracy for that call. This allows the method to identify

20
25
30

individual samples with low accuracy and flag them for follow-up. This is expected to nearly eliminate missed calls, especially at the early stages of pregnancy when fetal fractions are typically low. The presumption is that a no call is much preferred to a missed call, since a no call simply requires a redraw and reanalysis.

5

Converting Calculated Accuracies to Traditional Risk Scores

This method can offer an adjusted risk of aneuploidy for high-risk pregnant women, where the adjusted risk takes into account an *a priori* risk (Benn P, Cuckle H, Pergament E. Non-invasive prenatal diagnosis for Down syndrome: the paradigm will shift, but slowly. Ultrasound Obstet Gynecol 2012;39:127-130, which is hereby incorporated by reference in its entirety). Although the present method offers each patient a customized calculated accuracy, for clinical use these accuracies can be converted to traditional risk scores, which also denote the risk of an aneuploid pregnancy but are expressed as fractions. Traditional risk scores take into account various parameters, including maternal age-related risk and serum levels of biochemical markers, to offer a risk score above which a mother is considered high-risk and for whom follow-up invasive diagnostic procedures are recommended. This method significantly refines this risk score, thus reducing both the false positive and false negative rates, and offering a more accurate assessment of individual maternal risk. A calculated accuracy as used here is the likelihood that the ploidy call is correct, and is expressed as a percentage, but the calculated accuracies used in Experiment 19 do not include an age-related risk. Because calculation of a risk score typically includes an age-related risk, the calculated accuracies and traditional risk scores are not interchangeable; they must be combined to convert into a traditional risk score. The formula to combine the age-related risk with the calculated accuracy is:

25
$$\frac{R_1 R_2}{R_1 R_2 + [1 - R_1][1 - R_2]}$$

where R_1 is the risk score as calculated by the present method and R_2 is the risk score as calculated by first trimester screening.

SNP-Based Methods Negate Issues with Amplification Variation

30 An inherent drawback to the counting methods used by some other methods is that they determine fetal ploidy state by measuring the ratio of the number of reads mapping to the

chromosome of interest (*e.g.*, chromosome 21) to those mapping to a reference chromosome. Chromosomes with high or low GC content, including chromosomes 13, X, and Y, amplify with high variability. This can result in signal variation that is comparable in magnitude to the fetal cfDNA signal, which can confound copy number calls by altering the ratio of allele reads from the chromosome-of-interest to those from the reference chromosome. This can result in low accuracy for chromosomes 13, X, and Y. Significantly, this problem is exacerbated at low fetal cfDNA fractions, as tends to be the case at early gestational ages.

In contrast, SNP-based methods do not rely on consistent amplification levels between chromosomes, and are thus expected to provide results that are equally accurate across all chromosomes. Because the present method looks, in part, at relative counts of different alleles at polymorphic loci, which by definition differ only by a single nucleotide, it does not require the use of reference chromosomes, and this obviates the problems with chromosome-to-chromosome amplification variation that are inherent to methods that rely on quantitating read counts. Unlike quantitative methods that require reference chromosomes that are euploid, the present method is expected to be able to detect triploidy as well as copy-number neutral anomalies like uniparental disomy.

The Importance of Early Detection

Significantly, the combined at-birth prevalence of sex chromosome aneuploidies is higher than that of the most common autosomal aneuploidies (**Figure 32**). However, there are currently no routine non-invasive screening methods that reliably detect sex chromosome abnormalities. Thus, sex chromosome anomalies are generally detected prenatally as a side-effect of routine testing for Down syndrome or other autosomal aneuploidies; a large proportion of cases are missed entirely. Early and accurate detection is crucial for many of these disorders where early therapeutic intervention improves clinical outcomes. For example, Turner syndrome is often not diagnosed until adolescence, although its overall at-birth prevalence is 1 in 2,500 females. Growth hormone therapy is known to prevent short stature that results from the disorder, but treatments are significantly more effective when initiated prior to the age of 4. Additionally, estrogen replacement therapy can stimulate secondary sexual characteristics in patients with Turner syndrome, but again therapy must be initiated in the pre-teen years, before the syndrome is usually detected. Together, this underscores the importance of early, routine, and safe detection of sex chromosome aneuploidies. This method

offers the first approach with the potential to serve as a routine screen for sex chromosome anomalies.

Additional Applications

5 Because this method utilizes targeted amplification, it is uniquely poised to detect sub-microscopic anomalies, such as microdeletions and microduplications. Although non-targeted methods like MPSS have been shown to detect the DiGeorge microdeletion syndrome, this required a sufficiently high level of genomic coverage so as to make the approach unfeasible. This is because non-targeted amplification will be several orders of magnitude less efficient on
10 submicroscopic regions, as very small fraction of the sequencing reads will be informative. Additionally, the fact that the currently available methods have trouble accurately identifying ploidy state for the sex chromosomes suggests that they will also encounter variable amplification problems on smaller chromosomal segments.

 Similarly, SNP based methods can detect UPD disorders, which are copy number-
15 neutral anomalies that will not be detected by either the current noninvasive methods that rely on counting or the traditional invasive methods like amniocentesis and CVS that rely on cytogenetic karyotyping and/or fluorescence *in situ* hybridization. This is because SNP-based methods are uniquely able to distinguish individual haplotypes, whereas the clinically available MPSS-based and targeted methods amplify non-polymorphic loci and are thus unable to
20 determine, for example, whether the chromosomes-of-interest originate from the same parent. This means that these microdeletion/microduplication and UPD syndromes, including Prader-Willi, Angelman, and Beckwith-Wiedemann syndromes, are generally not diagnosed prenatally, and are often initially misdiagnosed postnatally. This significantly delays therapeutic intervention. Additionally, because this method targets SNPs, this method will also
25 facilitate parental haplotype reconstruction, allowing for detection of fetal inheritance of individual disease-linked loci (Kitzman JO, Snyder MW, Ventura M, *et al.* Noninvasive whole-genome sequencing of a human fetus. *Sci Transl Med* 2012;4:137ra76, which is hereby incorporated by reference in its entirety).

 The results presented here confirm the expanded scope of this method for identifying
30 prenatal aneuploidy. Specifically, by amplifying and sequencing 19,488 SNPs, this method is able to determine copy number at chromosomes 13, 18, 21, X, and Y, and is uniquely expected to detect other chromosomal abnormalities, such as triploidy and UPD, that are not detected by

any other clinically available non-invasive method. The increased clinical coverage and powerful sample-specific calculated accuracies suggest that this method may offer a viable adjunct to invasive testing for detecting fetal chromosomal aneuploidies.

5 All patents, patent applications, and published references cited herein are hereby incorporated by reference in their entirety. While the methods of the present disclosure have been described in connection with the specific embodiments thereof, it will be understood that it is capable of further modification. Furthermore, this application is intended to cover any variations, uses, or adaptations of the methods of the present disclosure, including such departures from the present disclosure as come within known or customary practice in the art
10 to which the methods of the present disclosure pertain, and as fall within the scope of the appended claims. For example, any of the methods disclosed herein for DNA can be readily adapted for RNA by including a reverse transcription step to convert the RNA into DNA. Examples that use polymorphic loci for illustration can be readily adapted for the amplification of nonpolymorphic loci if desired.

15

CLAIMS

What is claimed is:

1. A method for determining a ploidy status of a chromosome or chromosome
5 segment of a cancer in a host, the method comprising:
measuring the DNA in a mixed sample that comprises DNA from normal cells from the
host and DNA from the cancer at a plurality of at least 10 polymorphic loci on the
chromosome or chromosome segment;
determining allele counts at the plurality of polymorphic loci from the DNA
10 measurements made on the sample;
fitting, on a computer, the determined allele counts to expected allele counts at the
plurality of polymorphic loci for each of a plurality of ploidy hypotheses each specifying a
different possible ploidy state of the chromosome or chromosome segment for the cancer,
using a joint distribution model that takes into account the expected linkage between the
15 plurality of polymorphic loci on the chromosome or chromosome segment, to determine a
relative probability of each of the ploidy hypotheses; and
outputting the hypothesis with the greatest probability of being true as an indication of
the ploidy status of the chromosome or chromosome segment of the cancer.
2. The method of claim 1, wherein the sample is a plasma sample of the host.
- 20 3. The method of claim 1, wherein the method further comprises preferentially
enriching the DNA in the sample at a plurality of polymorphic loci prior to the measuring step.
4. The method of claim 3, wherein the preferentially enriching the DNA in the
sample at the plurality of polymorphic loci comprises:
obtaining a plurality of pre-circularized probes in which each probe targets one of the
25 polymorphic loci, wherein the 3' and 5' end of each probe is designed to hybridize to a region
of DNA that is separated from the polymorphic site of the locus by a small number of bases,
and wherein the small number is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19,
20, 21 to 25, 26 to 30, 31 to 60, or a combination thereof;
hybridizing the pre-circularized probes to DNA from the sample;
30 filling the gap between the hybridized probe ends using DNA polymerase; and
circularizing the pre-circularized probe; and amplifying the circularized probe.

5. The method of claim 3, wherein the preferentially enriching the DNA at the plurality of polymorphic loci comprises:

obtaining a plurality of ligation-mediated PCR probes in which each PCR probe targets one of the polymorphic loci, wherein the corresponding upstream and downstream PCR probes are designed to hybridize to a region of DNA on one strand of DNA that is separated from the polymorphic site of the locus by a small number of bases, and wherein the small number is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 to 25, 26 to 30, 31 to 60, or a combination thereof;

hybridizing the ligation-mediated PCR probes to the DNA from the sample;

filing the gap between the ligation-mediated PCR probe ends using DNA polymerase;

ligating the ligation-mediated PCR probes; and

amplifying the ligated ligation-mediated PCR probes.

6. The method of claim 3, wherein the preferentially enriching the DNA at the plurality of polymorphic loci comprises:

obtaining a plurality of hybrid capture probes that target the polymorphic loci;

hybridizing the hybrid capture probes to the DNA in the sample; and

physically removing some or all of the unhybridized DNA from the sample of DNA.

7. The method of claim 3, wherein the preferentially enriching the DNA at the plurality of polymorphic loci comprises:

obtaining a plurality of inner forward primers in which each primer targets one of the polymorphic loci, wherein the 3' end of each of the inner forward primers is designed to hybridize to a region of DNA upstream from the polymorphic site of the locus and separated from the polymorphic site by a small number of bases, and wherein the small number is selected from the group consisting of 1, 2, 3, 4, 5, 6 to 10, 11 to 15, 16 to 20, 21 to 25, 26 to 30, and 31 to 60 base pairs;

optionally obtaining a plurality of inner reverse primers in which each primer targets one of the polymorphic loci, wherein the 3' end of each of the inner reverse primers is designed to hybridize to a region of DNA upstream from the polymorphic site of the locus and separated from the polymorphic site by a small number of bases, and wherein the small number is selected from the group consisting of 1, 2, 3, 4, 5, 6 to 10, 11 to 15, 16 to 20, 21 to 25, 26 to 30, and 31 to 60 base pairs;

hybridizing the inner primers to the DNA; and

amplifying the DNA using the polymerase chain reaction to form amplicons.

8. The method of claim 7, further comprising:

obtaining a plurality of outer forward primers in which each primer targets one of the polymorphic loci, wherein each of the outer forward primers is designed to hybridize to the region of DNA upstream from the corresponding inner forward primer;

optionally obtaining a plurality of outer reverse primers in which each primer targets one of the polymorphic loci, wherein each of the outer reverse primers is designed to hybridize to the region of DNA immediately downstream from the corresponding inner reverse primer;

hybridizing the first primers to the DNA; and

amplifying the DNA using the polymerase chain reaction.

9. The method of claim 7, further comprising:

obtaining a plurality of outer reverse primers in which each primer targets one of the polymorphic loci, wherein each of the outer reverse primers is designed to hybridize to the region of DNA immediately downstream from the corresponding inner reverse primer;

optionally obtaining a plurality of outer forward primers in which each primer targets one of the polymorphic loci, wherein each of the outer forward primers is designed to hybridize to the region of DNA upstream from the corresponding inner forward primer;

hybridizing the first primers to the DNA; and

amplifying the DNA using the polymerase chain reaction.

10. The method of claim 7, further comprising:

appending universal adapters to the DNA in the sample; and

amplifying the DNA in the sample using the polymerase chain reaction.

11. The method of claim 7, wherein the amplifying the DNA is done in one or a plurality of individual reaction volumes, wherein each individual reaction volume contains more than 500 different forward and reverse primer pairs.

12. The method of claim 7, further comprises dividing the first sample into a plurality of portions, and wherein the DNA in each portion is preferentially enriched at a subset of the plurality of polymorphic loci.

13. The method of claim 7, wherein the inner primers are selected by identifying primer pairs likely to form undesired primer duplexes and removing from the plurality of primers at least one of the pair of primers identified as being likely to form undesired primer duplexes.

14. The method of claim 1, further comprising obtaining normal host cell genotypic data at the plurality of polymorphic loci from normal cells from the host.

15. The method of claim 14, wherein the step of fitting the allele counts to expected allele counts at each of the plurality of polymorphic loci for each of the plurality of ploidy
5 hypotheses, uses the normal host cell genotypic data.

16. The method of claim 14, wherein the sample is isolated from host plasma, and wherein the obtaining normal host cell genotypic data is done by estimating the normal cell genotypic data from the DNA measurements made on the sample.

17. The method of claim 3, wherein the preferential enrichment results in an
10 average degree of allelic bias between the sample after preferential enrichment and the sample prior to preferential enrichment of no more than a factor of 1.2.

18. The method of claim 1, wherein the plurality of polymorphic loci are single nucleotide polymorphisms.

19. The method of claim 1, wherein the step of measuring the DNA in the sample is
15 done by sequencing.

20. The method of claim 1, wherein the measurements of the DNA in the sample at the plurality of polymorphic loci are analyzed to estimate the haplotypes present on the chromosome or chromosome segment.

21. The method of claim 1, wherein the step of fitting is performed using the joint
20 distribution model that takes into account the probability of chromosomes crossing over at different locations in the chromosome or chromosome segment to model dependence between blocks of polymorphic alleles on the chromosome or chromosome segment.

22. The method of claim 1, wherein the step of fitting the allele counts to expected allele counts is done using a method that does not require the use of a reference chromosome.

25 23. The method of claim 1, wherein the step of fitting the allele counts to expected allele counts makes use of an estimated fraction of circulating cancer DNA in the sample.

24. The method of claim 1, wherein the step of fitting the allele counts to expected allele counts further comprises determining relative probabilities for each of the ploidy hypotheses using one or more statistical techniques selected from the group consisting of a
30 read count analysis, comparing heterozygosity rates, a statistic that is only available when normal host cell genetic information is used, the probability of normalized genotype signals for

certain normal host cell contexts, a statistic that is calculated using an estimated circulating cancer DNA fraction of the sample, and combinations thereof.

25. The method of claim 1, wherein a confidence estimate is calculated for the outputted hypothesis.

5 26. The method of claim 1, wherein the allele counts measured on the sample comprise quantitative measurements of the number of independent observations of each allele at each polymorphic locus, and wherein the expected allele counts comprise expected quantitative measurements of the number of independent observations of each allele at each polymorphic locus.

10 27. The method of claim 1, wherein determining the relative probability of each of the ploidy hypotheses does not involve calculating the ratio of the alleles counts measured on the sample at a polymorphic locus.

28. The method of claim 1, further comprising:

15 measuring the DNA in a normal sample from the host that comprises DNA from normal cells from the host and not from the cancer, at the plurality of at least 10 polymorphic loci on the chromosome or chromosome segment;

 determining, on a computer, normal host cell allele counts at the plurality of polymorphic loci from the DNA measurements made on the normal sample;

20 phasing the normal host cell allele counts taking into account the expected linkage of the polymorphic loci of the plurality of polymorphic loci, and

 using the phased normal host cell allele counts to determine the expected allele counts at each of the plurality of polymorphic loci.

ABSTRACT OF THE DISCLOSURE

5 The invention provides methods for simultaneously amplifying multiple nucleic acid
regions of interest in one reaction volume as well as methods for selecting a library of primers
for use in such amplification methods. The invention also provides library of primers with
desirable characteristics, such as minimal formation of amplified primer dimers or other non-
target amplicons.

10

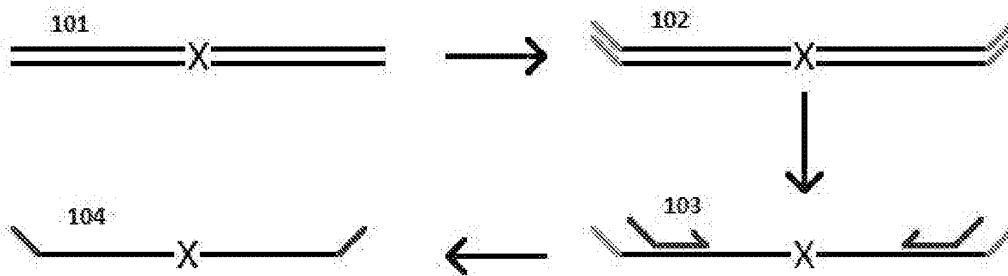


Figure 1

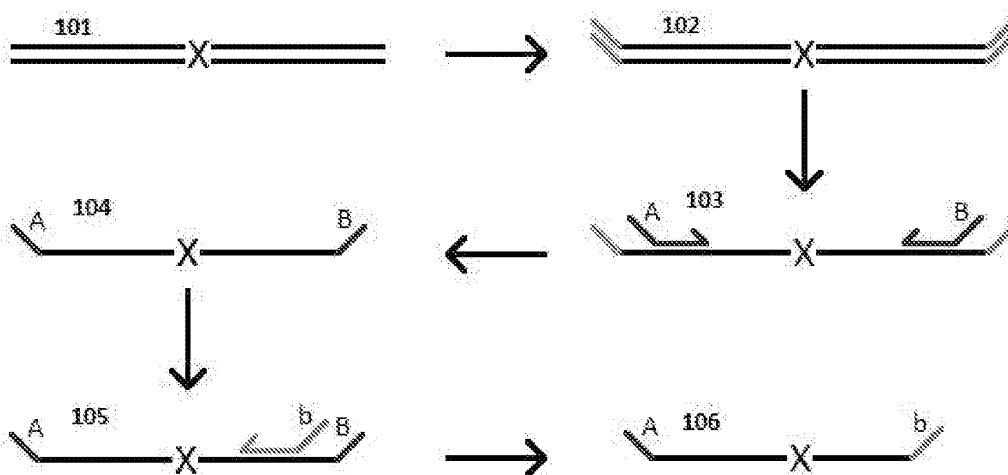


Figure 2

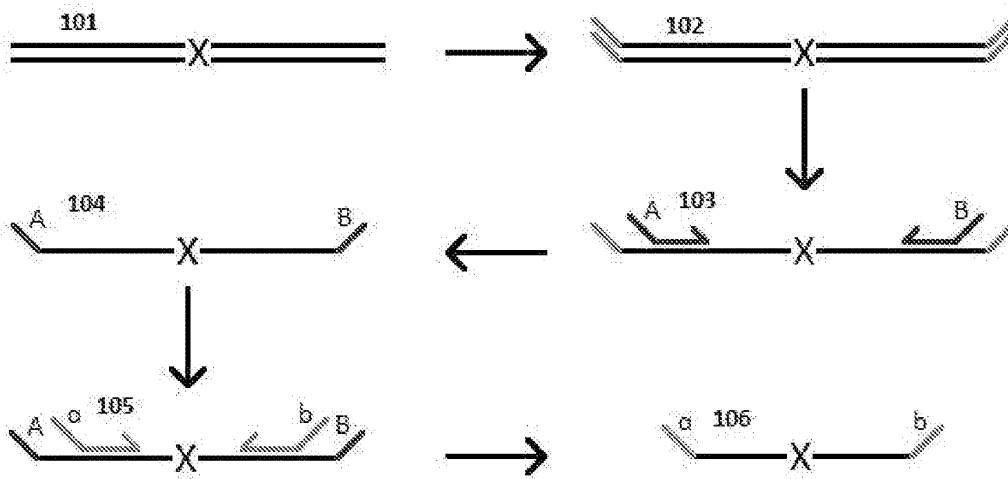


Figure 3

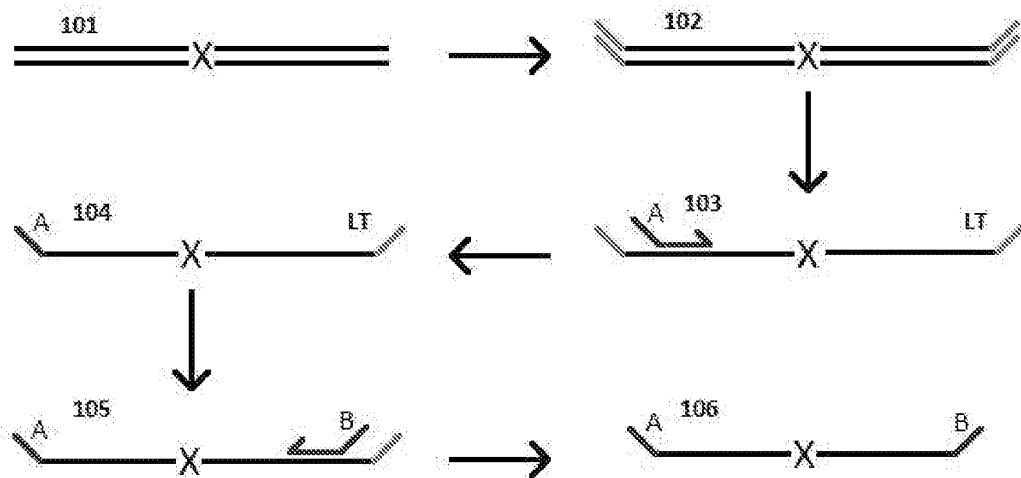


Figure 4

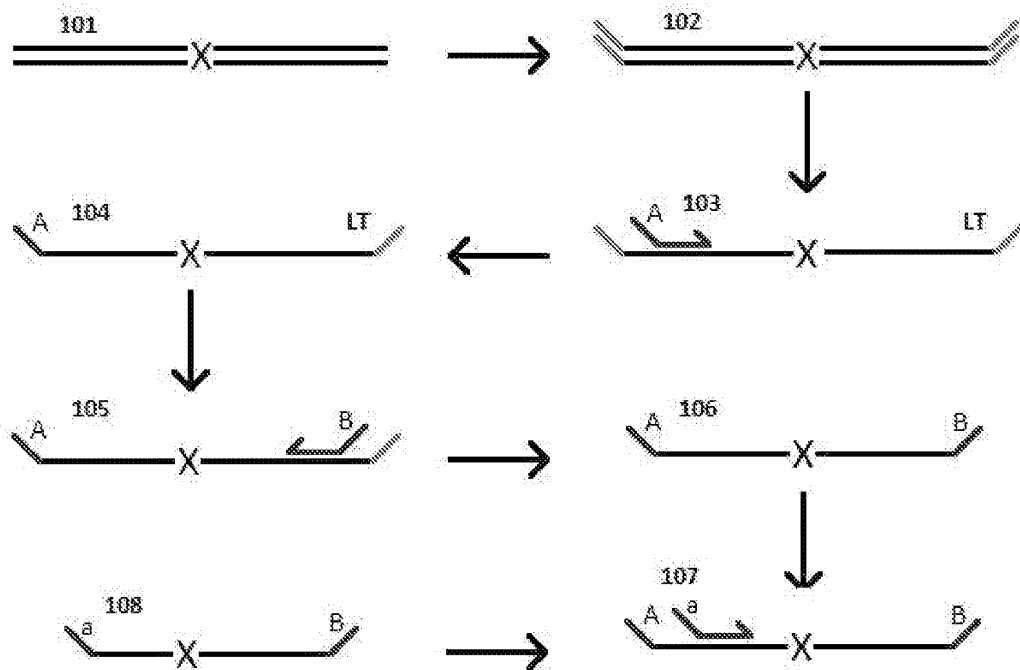


Figure 5

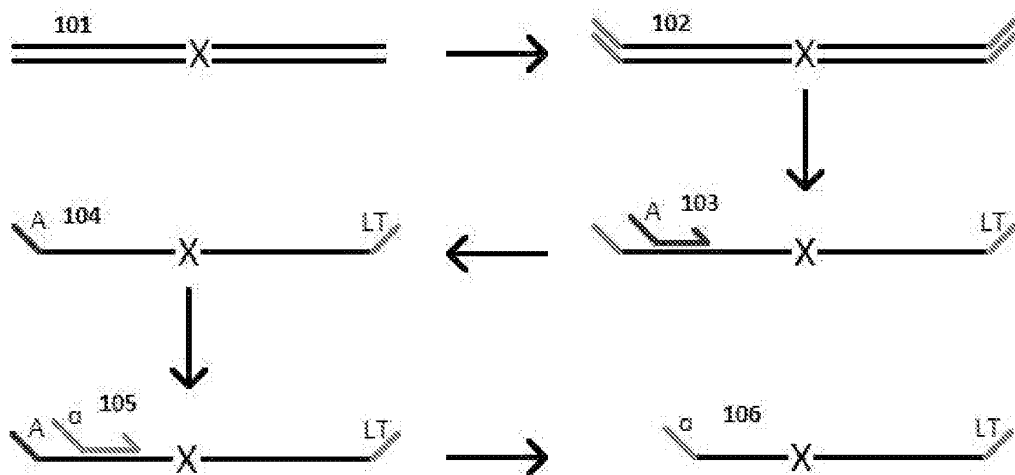


Figure 6

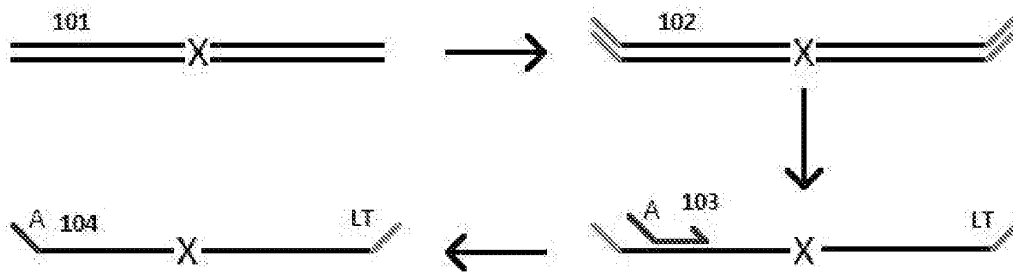


Figure 7

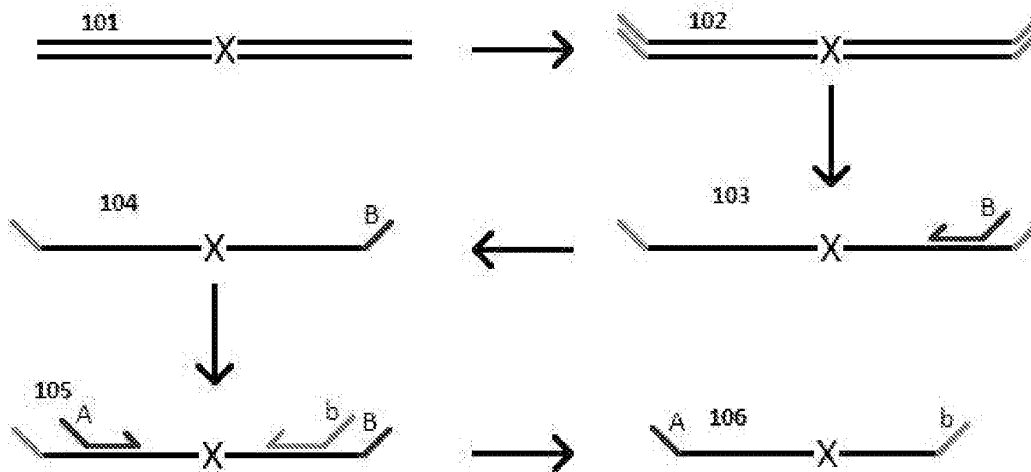


Figure 8

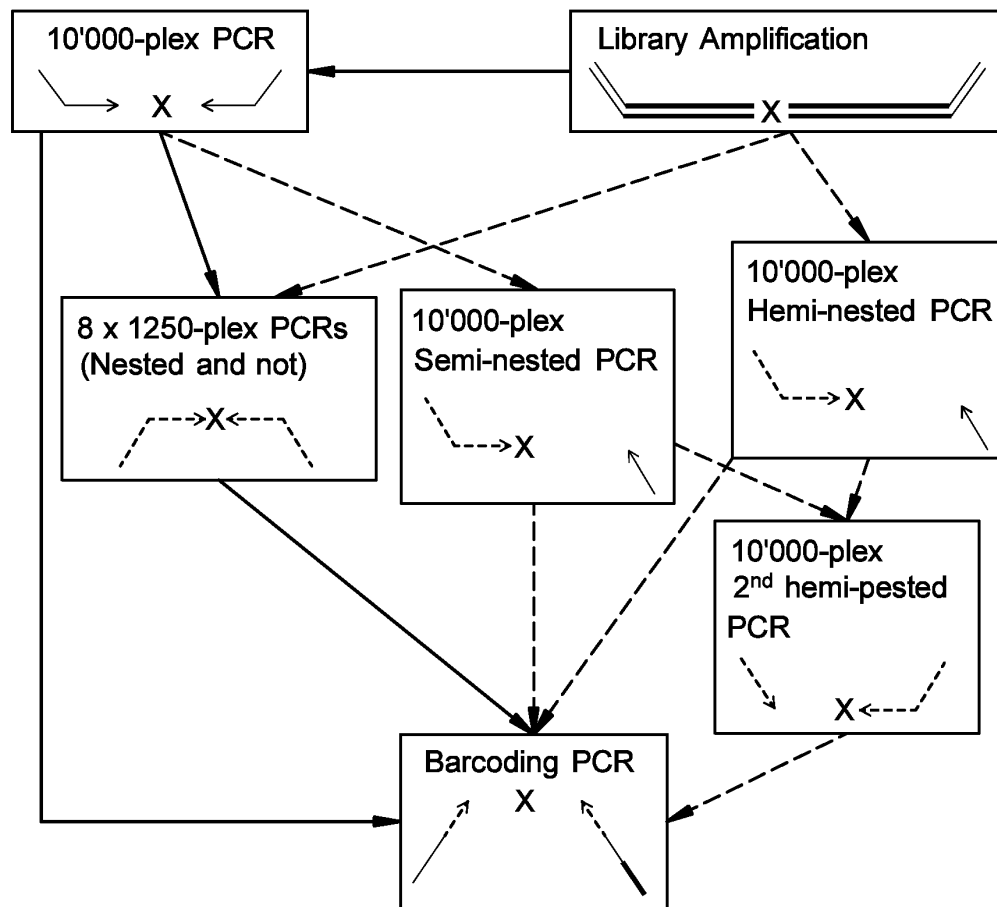


Figure 9

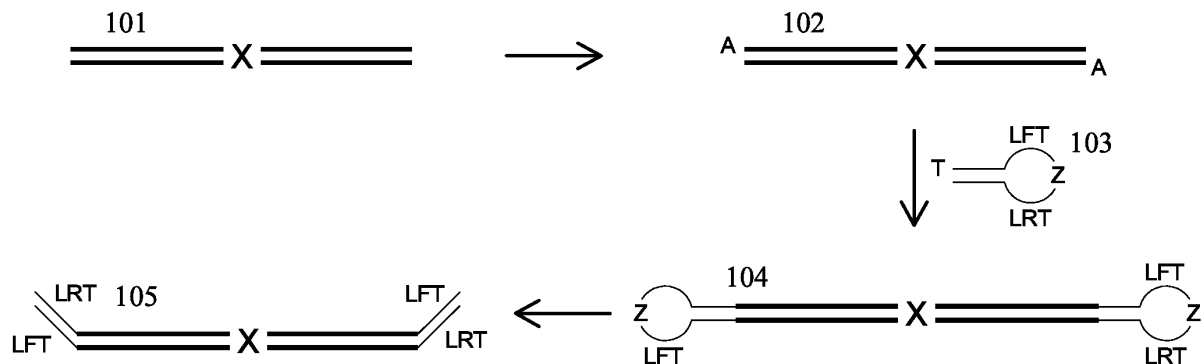


Figure 10

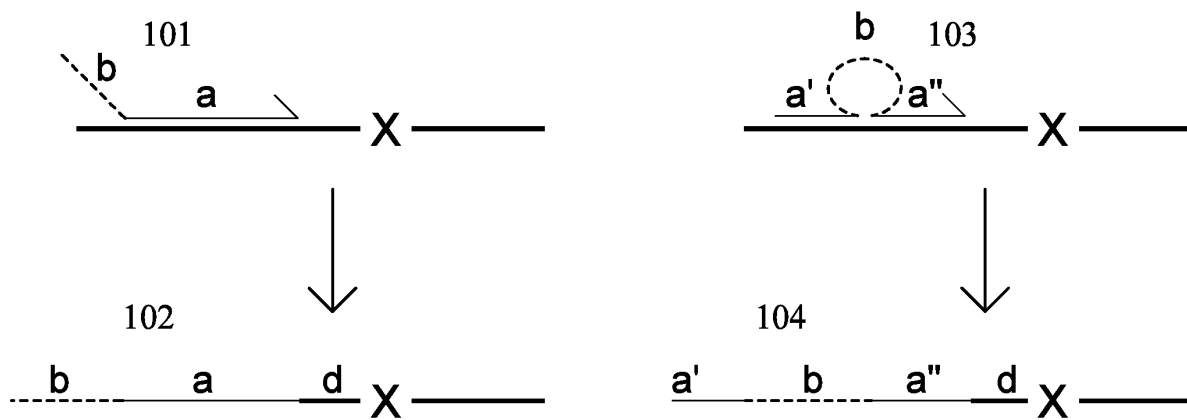


Figure 11

The sequencing-adaptor sequence is located inside the primer sequence and flanked by target specific sequence on both sides. 10 bases are target-specific at the 3'-end of each primer. Primers were tested successfully in real-time PCR. For sequencing this reduces the number of primer bases that need to be sequenced.

seqid1	seqid2	seqid3	seqid4	seqid5	seqid6
rs8130564	int-tag 1.10	AACTCACATAGC <u>ACACGACGC</u> TCTTCCGATCT TGCAAGCAC			
rs2832093	int-tag 2.10	TCCTCTGTG <u>ACACGACGC</u> TCTTCCGATCTCCCTGCTCT			
rs12011281	int-tag 3.10	tctctctctACACGACGCTCTTCCGATCTcGGGCTGTCA			
rs6719561	int-tag 4.10	TACATCCTTGAGACACGACGCTCTTCCGATCTGCTGTGCAGT			
rs10187018	int-tag 5.10	tttgcttgagctACACGACGCTCTTCCGATCTcgggagtttc			
rs10460481	int-tag 6.10	gtcttatgttgACACGACGCTCTTCCGATCTcaagccagt			

The sequencing-adaptor sequence is located inside the primer sequence and flanked by target specific sequence on both sides. The internal tag is formed into a hairpin structure by 10 complementary bases on either end. This brings the target-specific ends of the primer into close proximity and hinders unspecific binding to the "internal tag". 10 bases are target-specific at the 3'-end of each primer. Primers were tested successfully in real-time PCR.

rs8130564	loop-int-tag 1.10	AACTCACATAGC <u>tgatcgg</u> ACACGACGCCTCTCCGATCTTGCAAGCACAC	seqid7
rs2832093	loop-int-tag 2.10	TCCTCTGTG <u>tgatcgg</u> ACACGACGCCTCTCCGATCTCCCTGCTCTT	seqid8
rs12011281	loop-int-tag 3.10	tcctctttgatcggTACACGACGCCTCTCCGATCTcGGGCTGTCA	seqid9
rs6719561	loop-int-tag 4.10	TACATCCTTGAGtgatcggTACACGACGCCTCTCCGATCTGCTGTCAGT	seqid10
rs10187018	loop-int-tag 5.10	tttgcttgagcttgatcggTACACGACGCCTCTCCGATCTcgggagtttc	seqid11
rs10460481	loop-int-tag 6.10	gtcttatgtggtgatcggTACACGACGCCTCTCCGATCTcaagccagt	seqid12

Figure 12

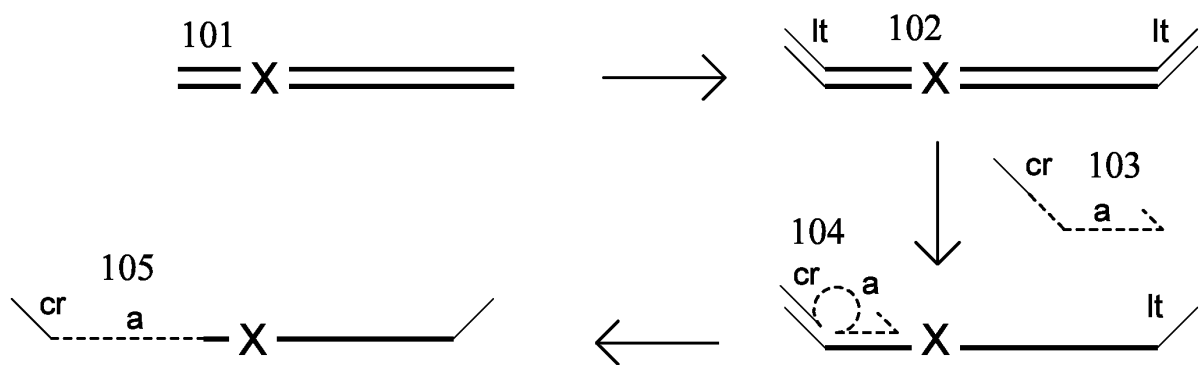


Figure 13

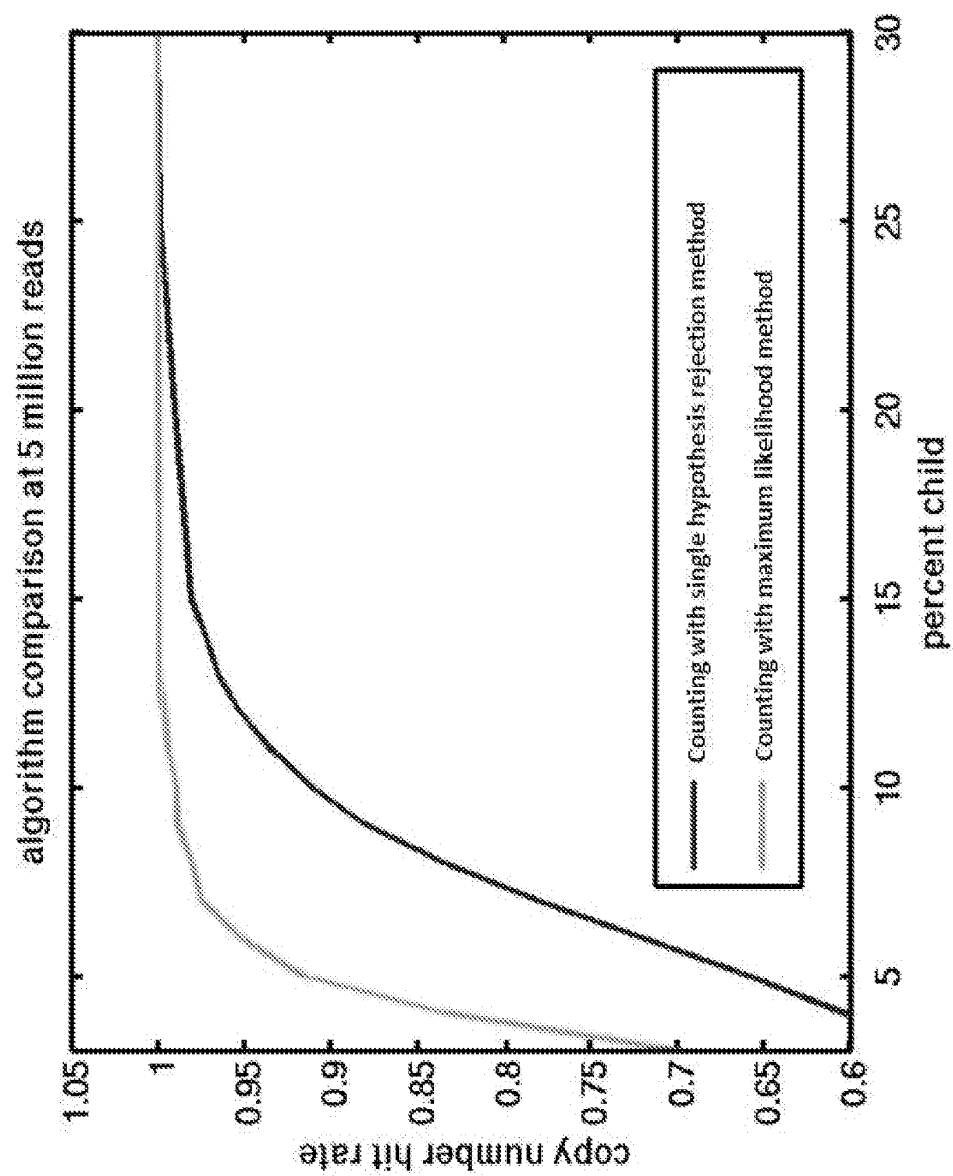


Figure 14

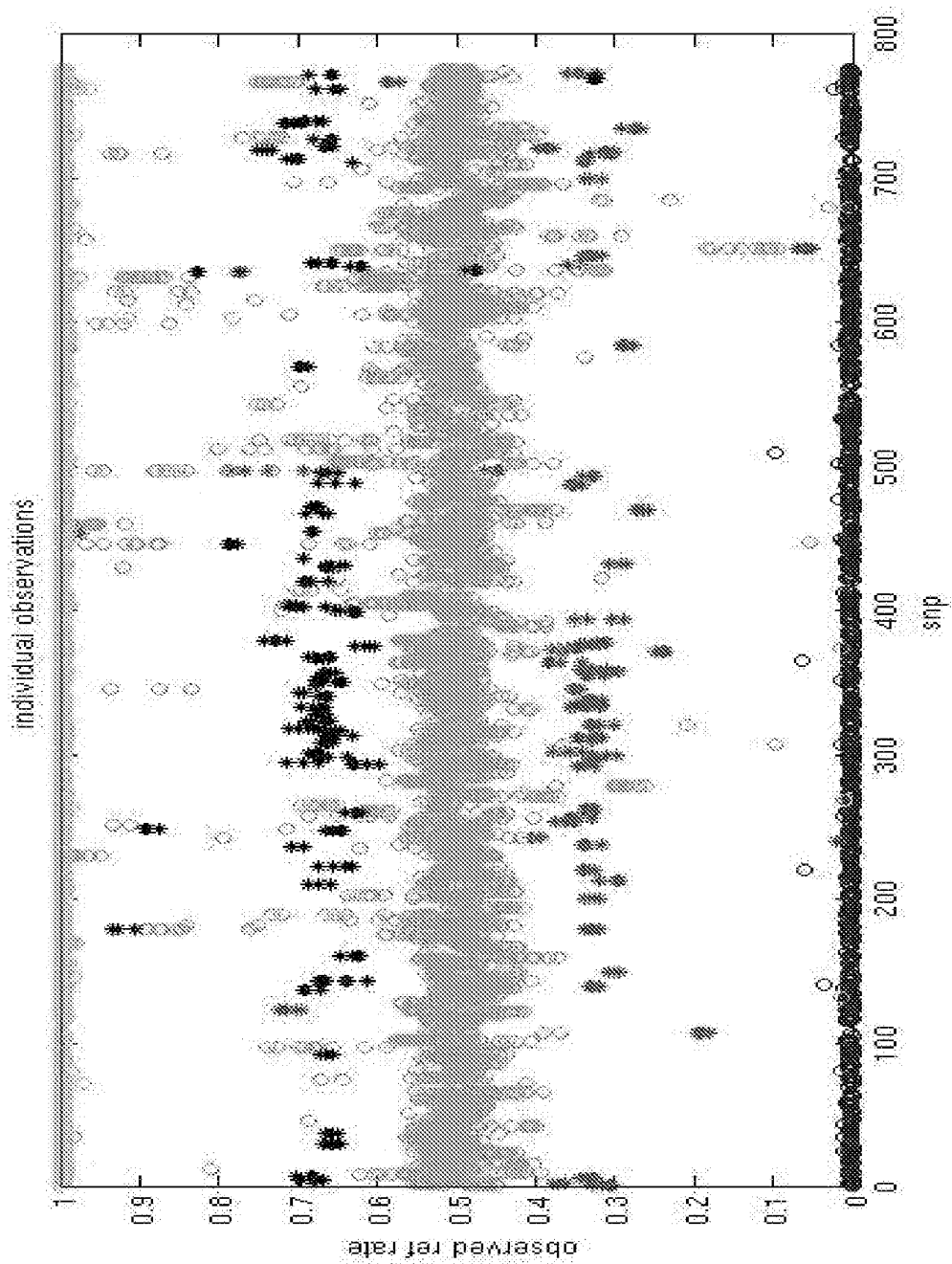


Figure 15

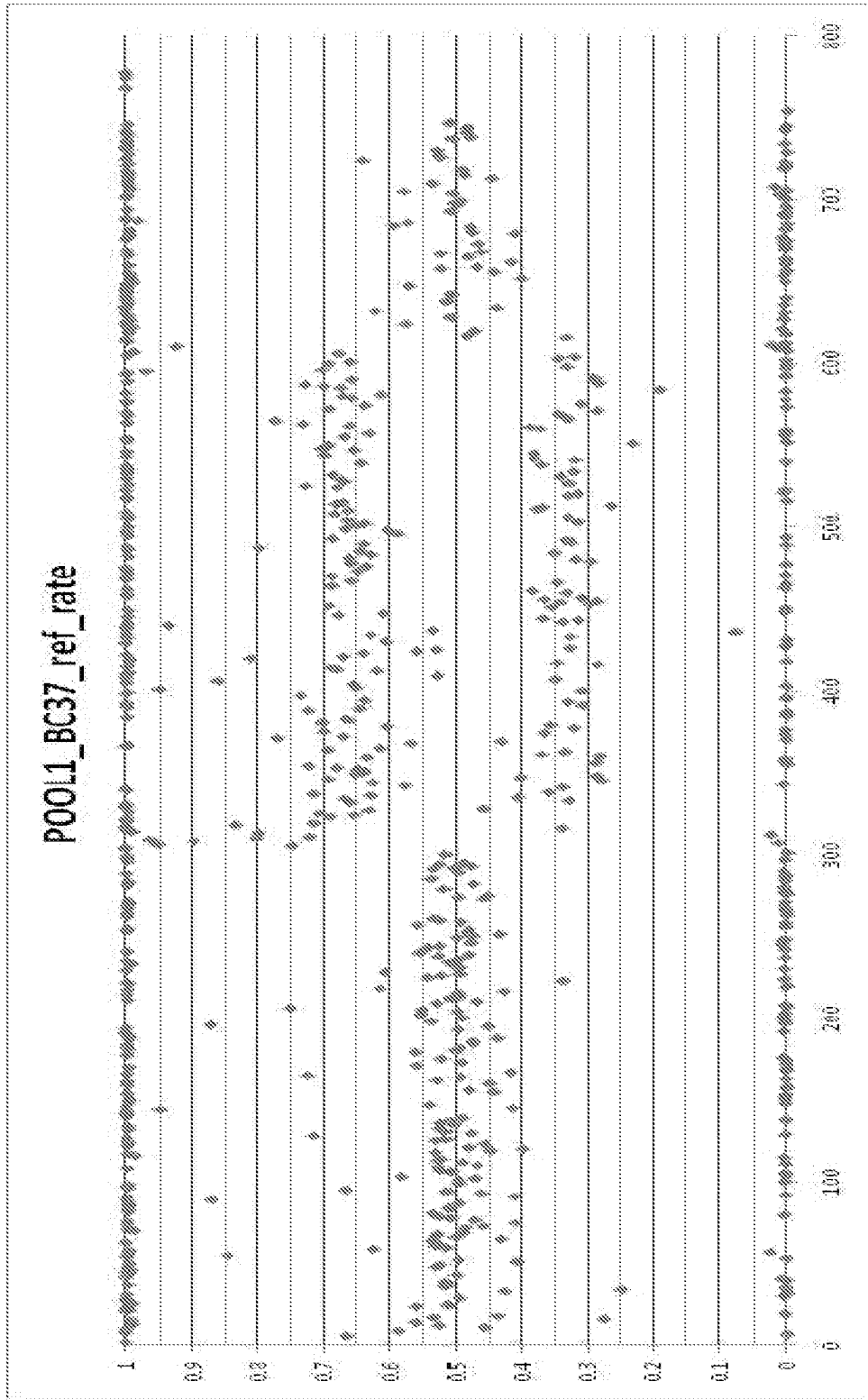


Figure 16

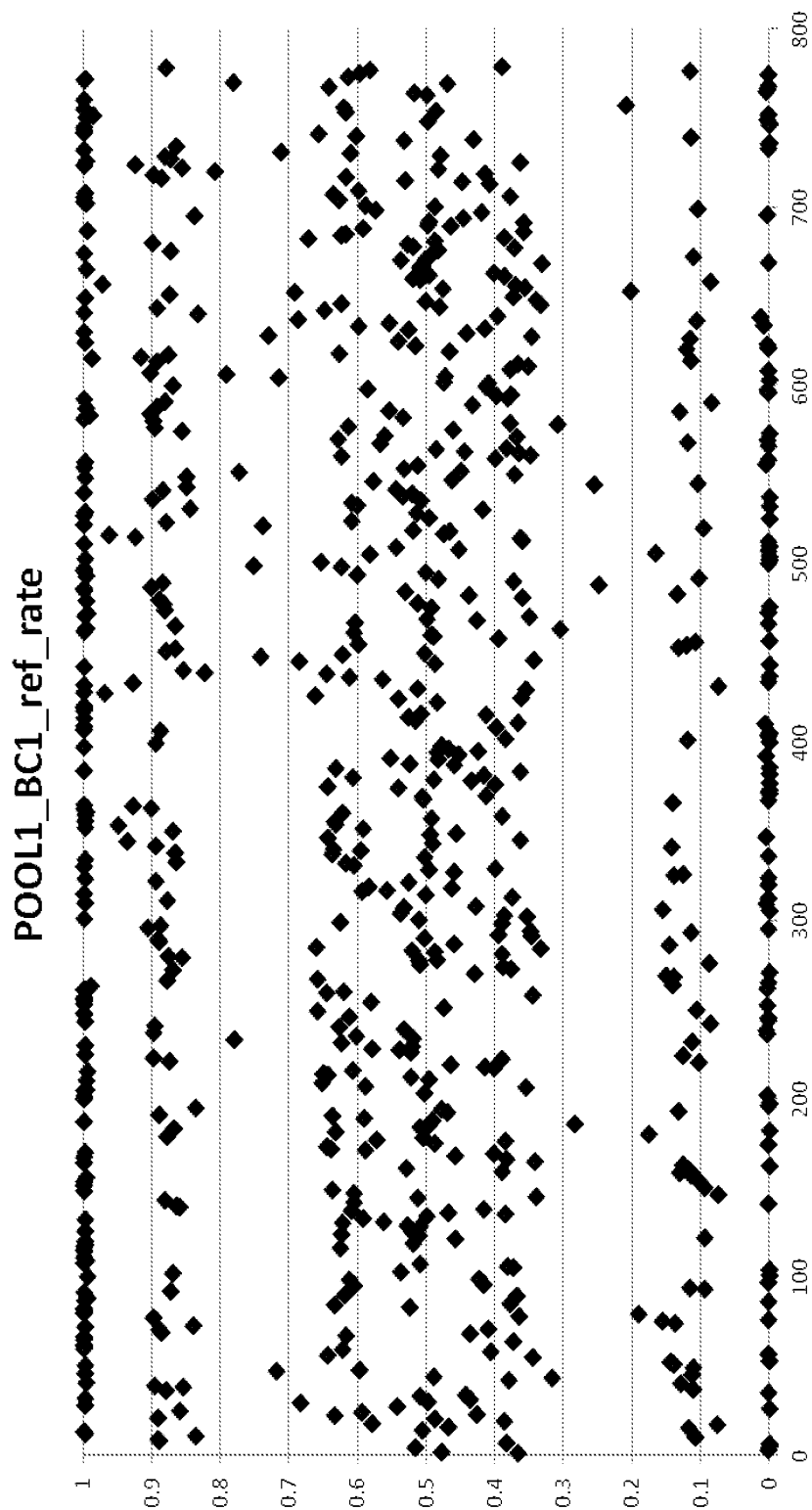


Figure 17A

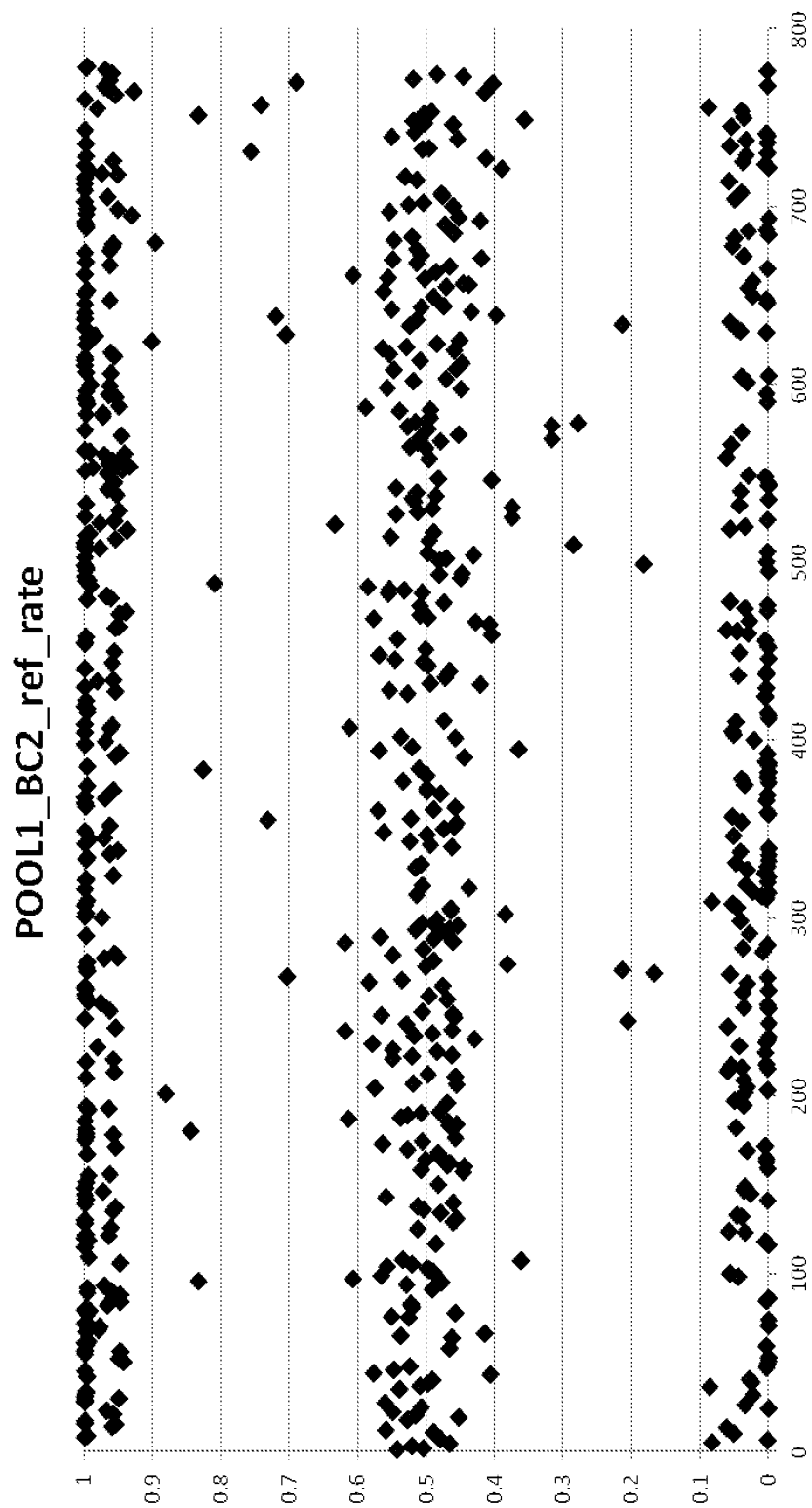


Figure 17B

No fetal DNA

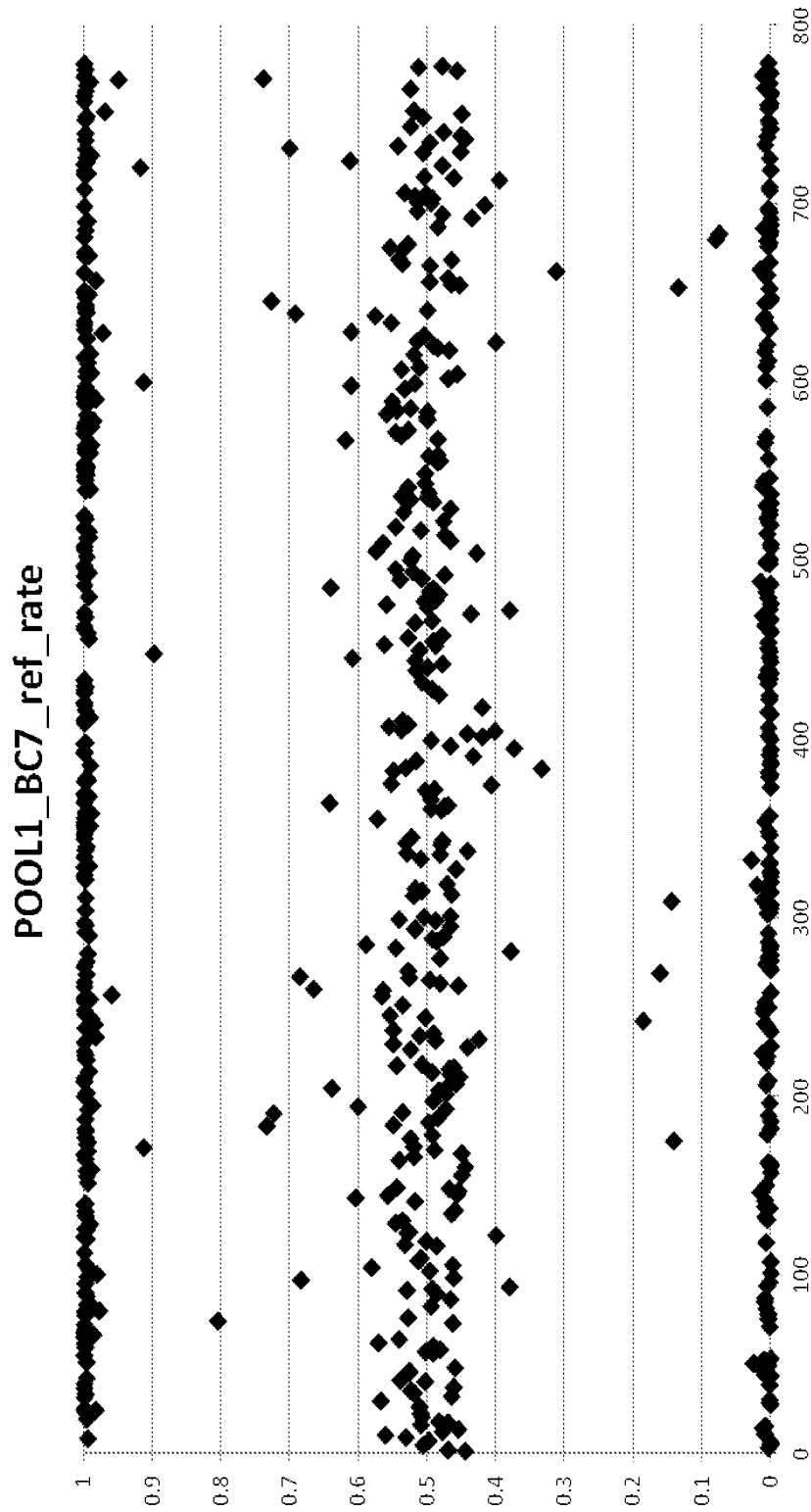


Figure 17C

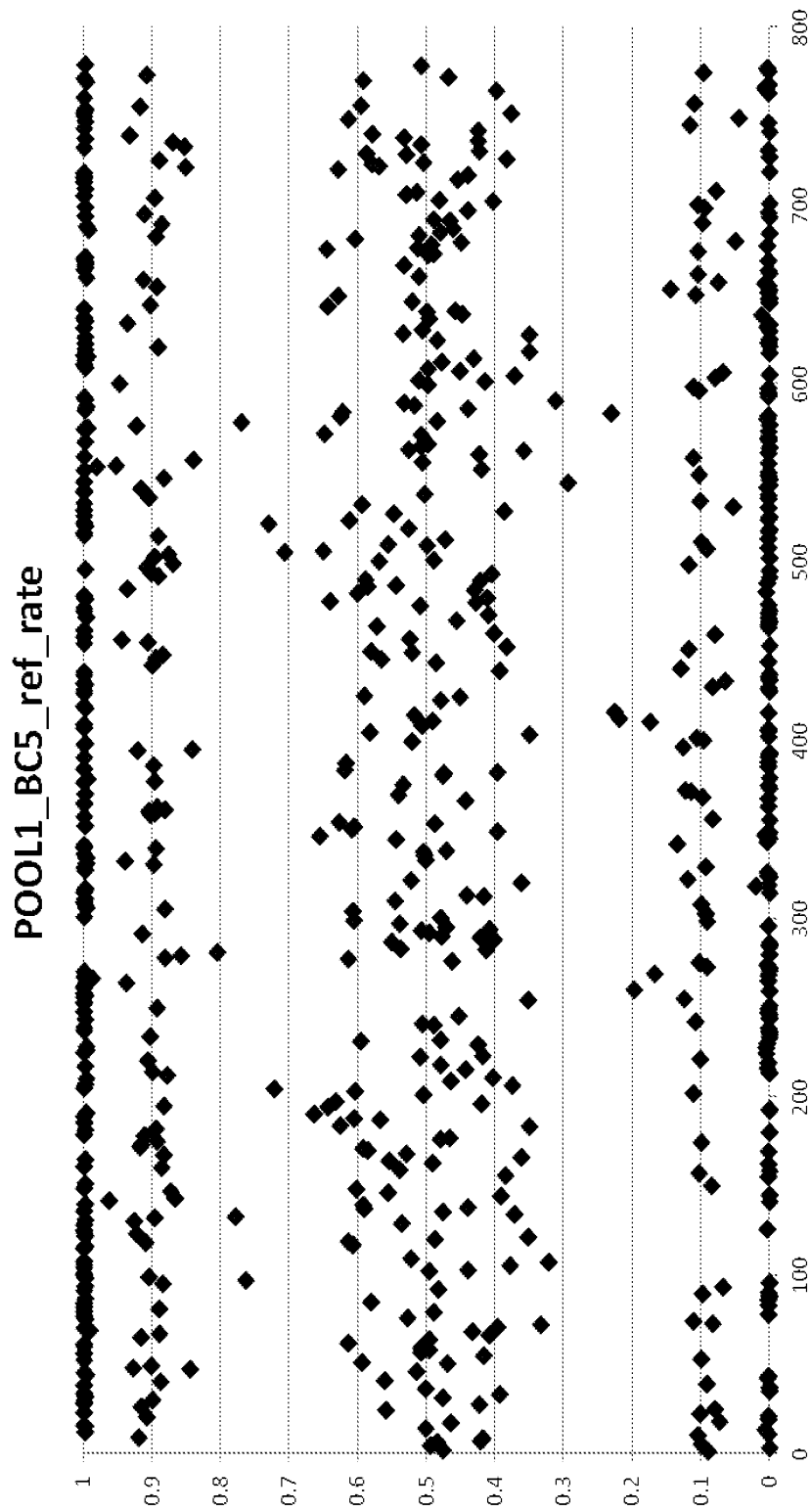


Figure 17D

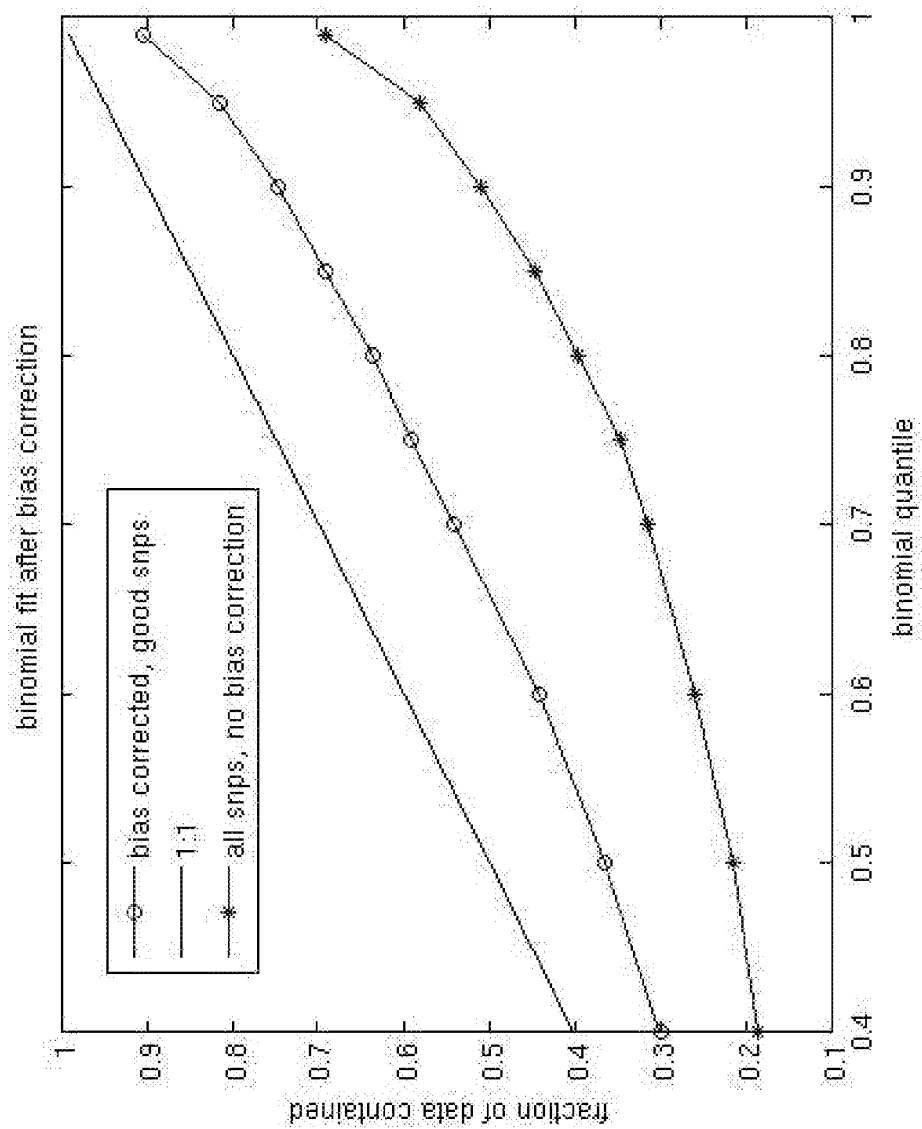


Figure 18

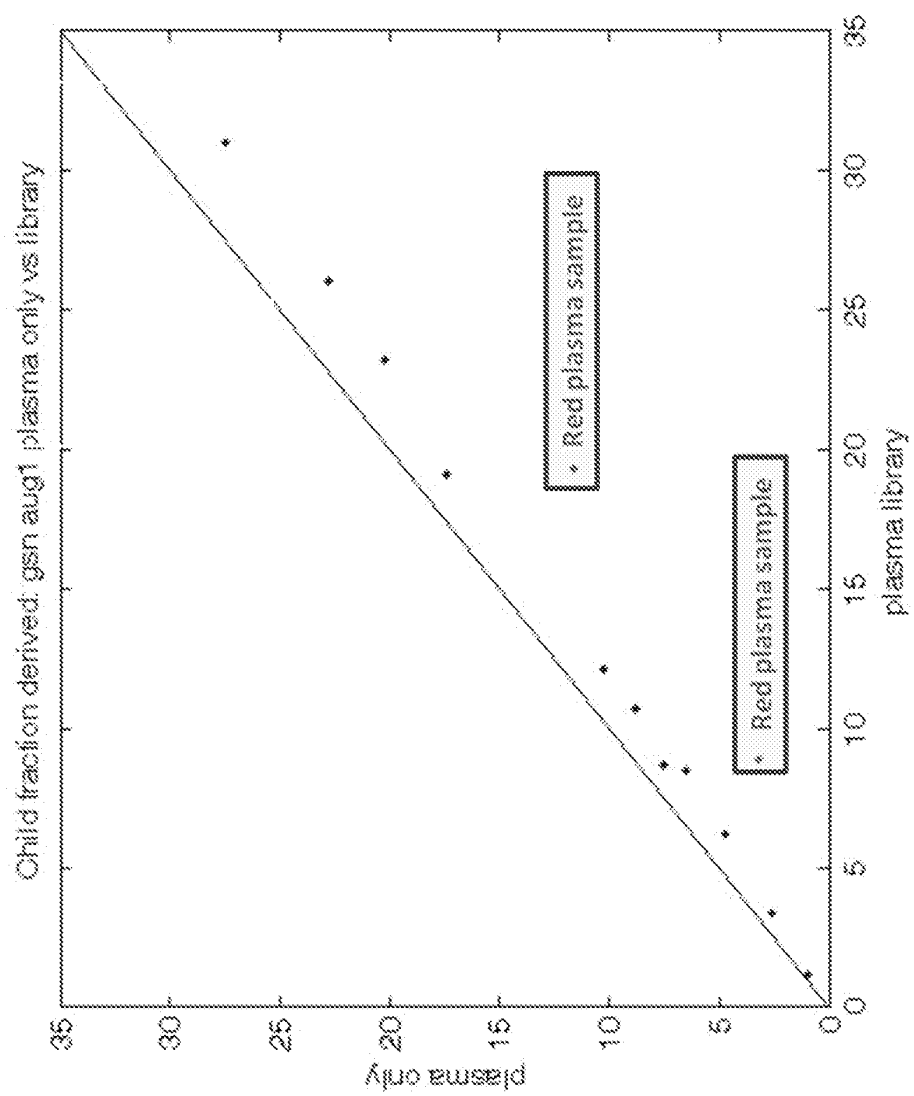


Figure 19

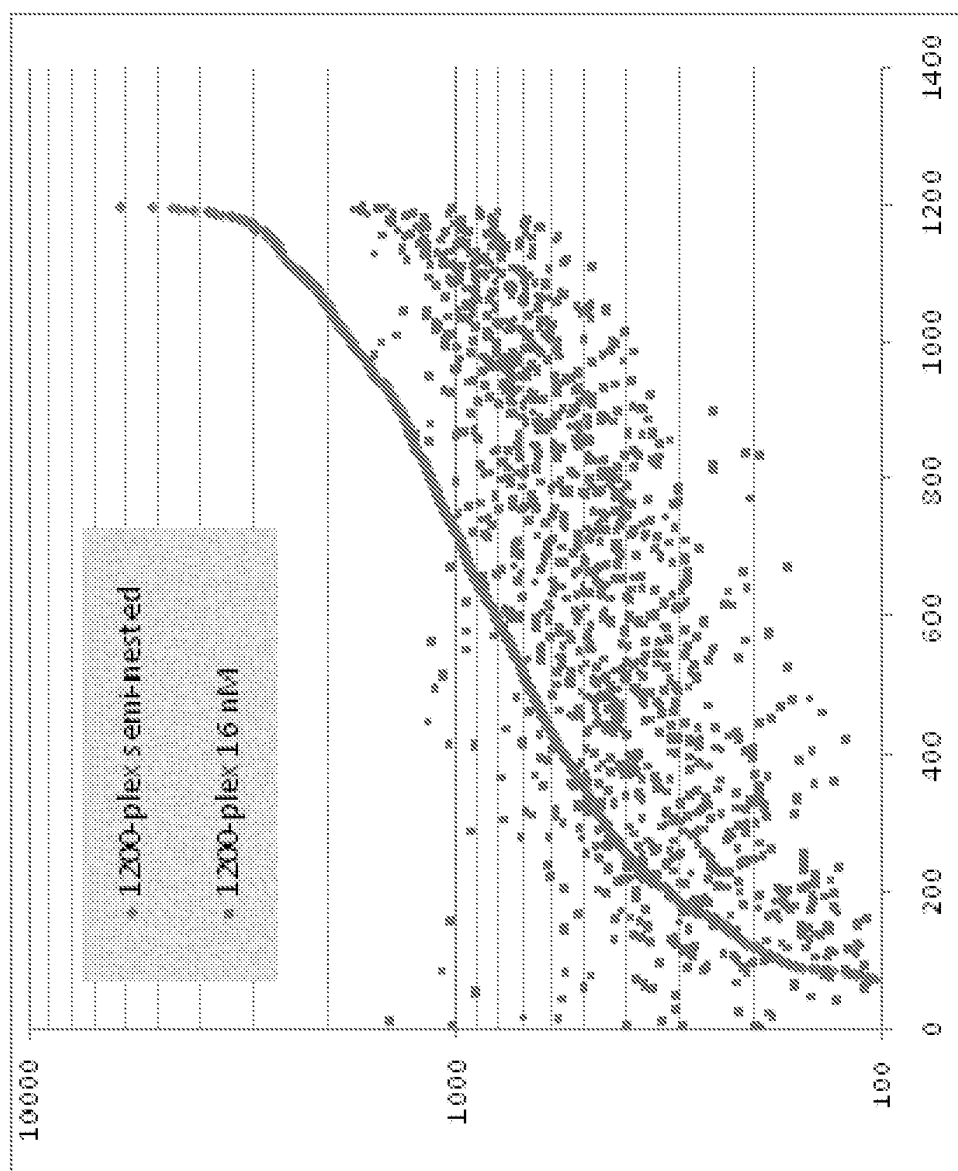


Figure 20

DOR range direct 1200-plex, 10 nM, 10 min TA

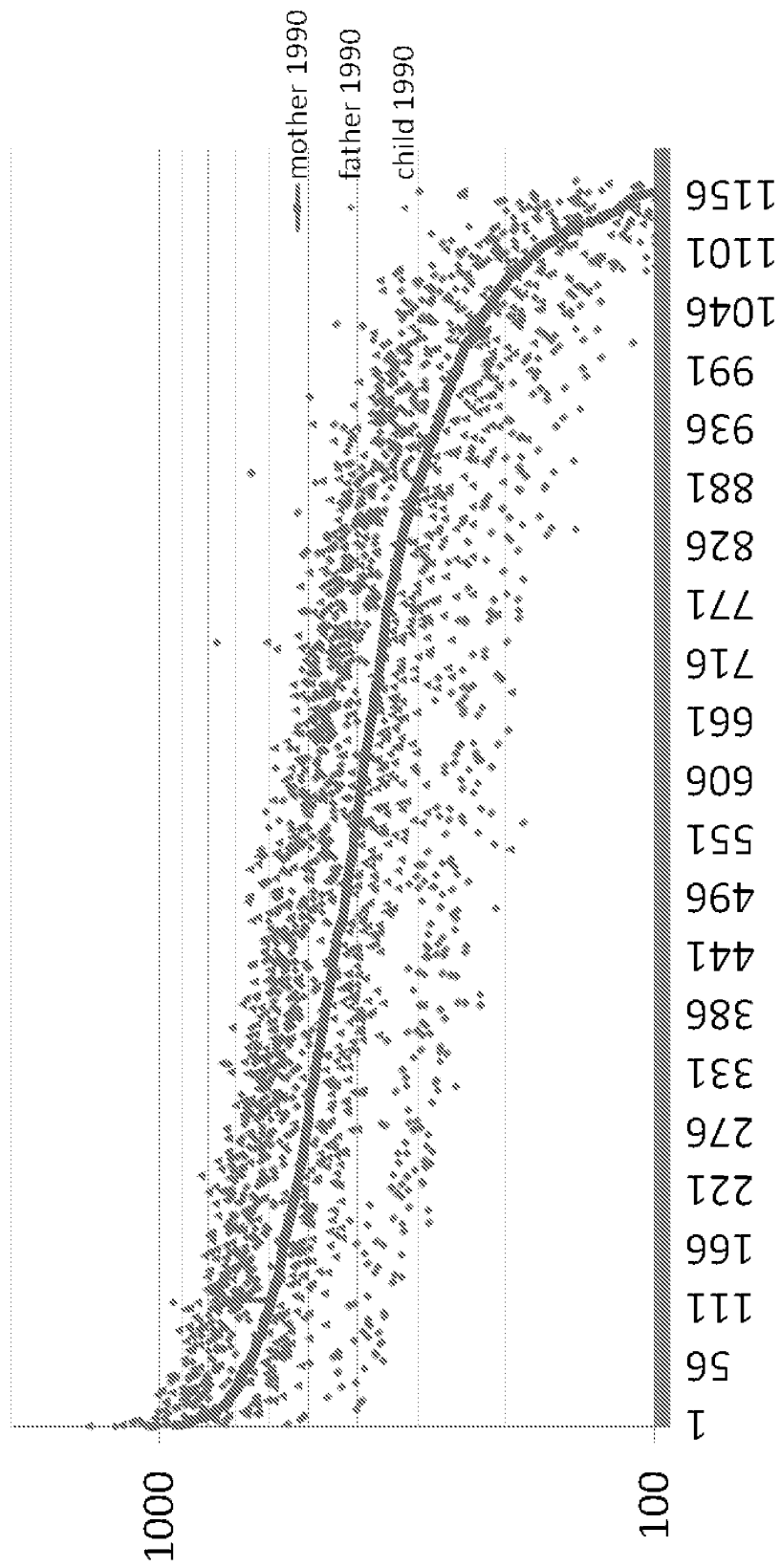


Figure 21

DOR range semi-nested 1200-plex, 10 nM, 10 min TA

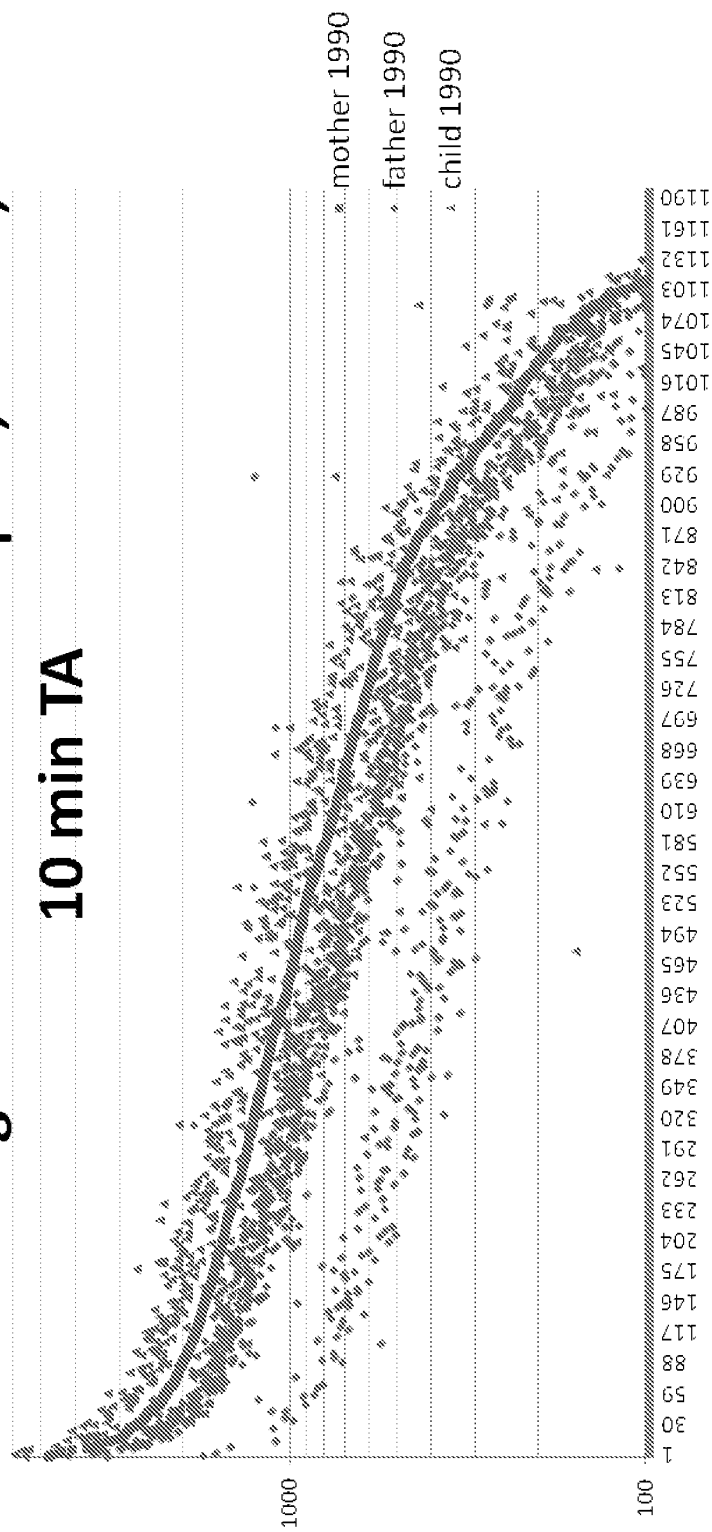


Figure 22

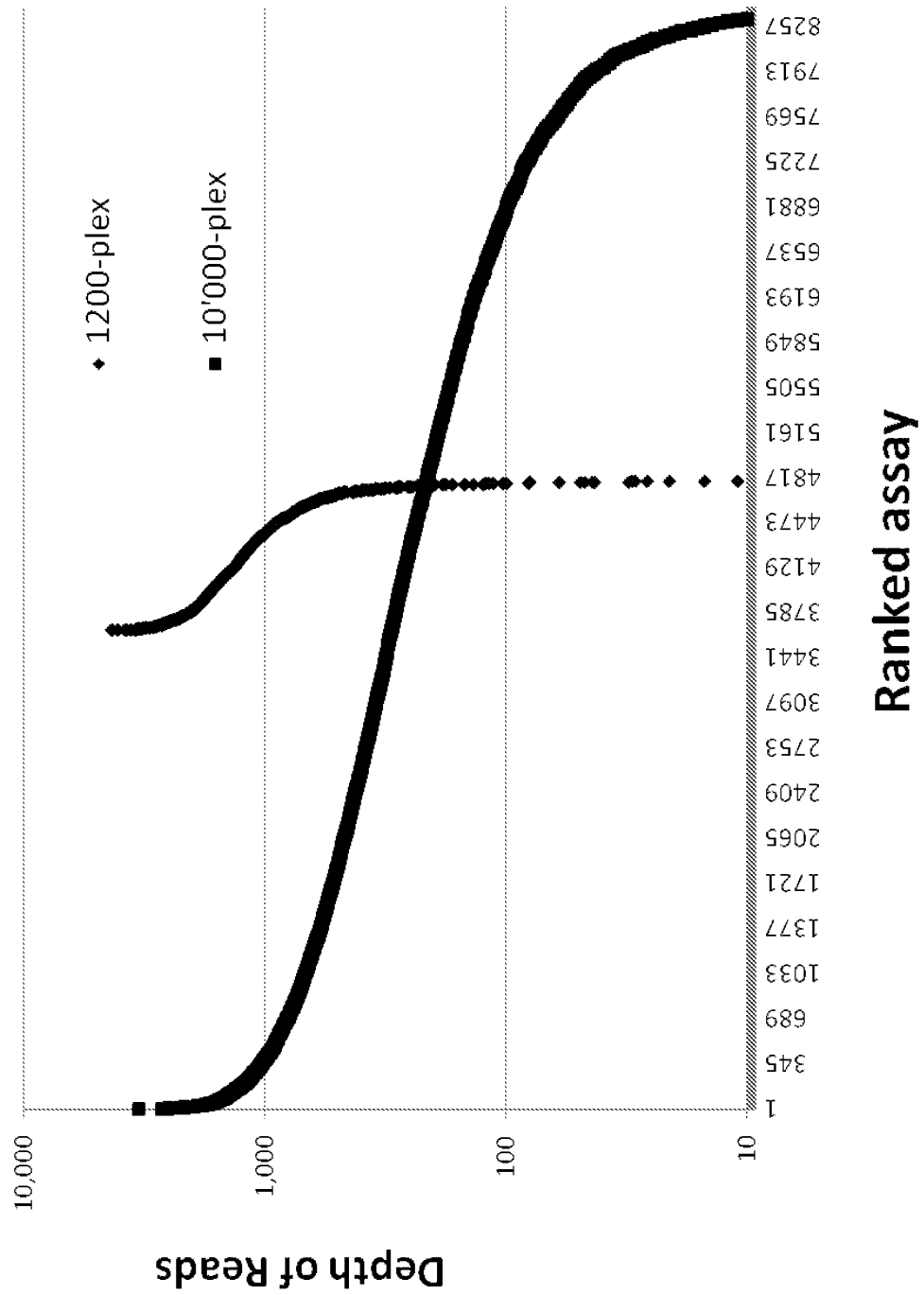


Figure 23

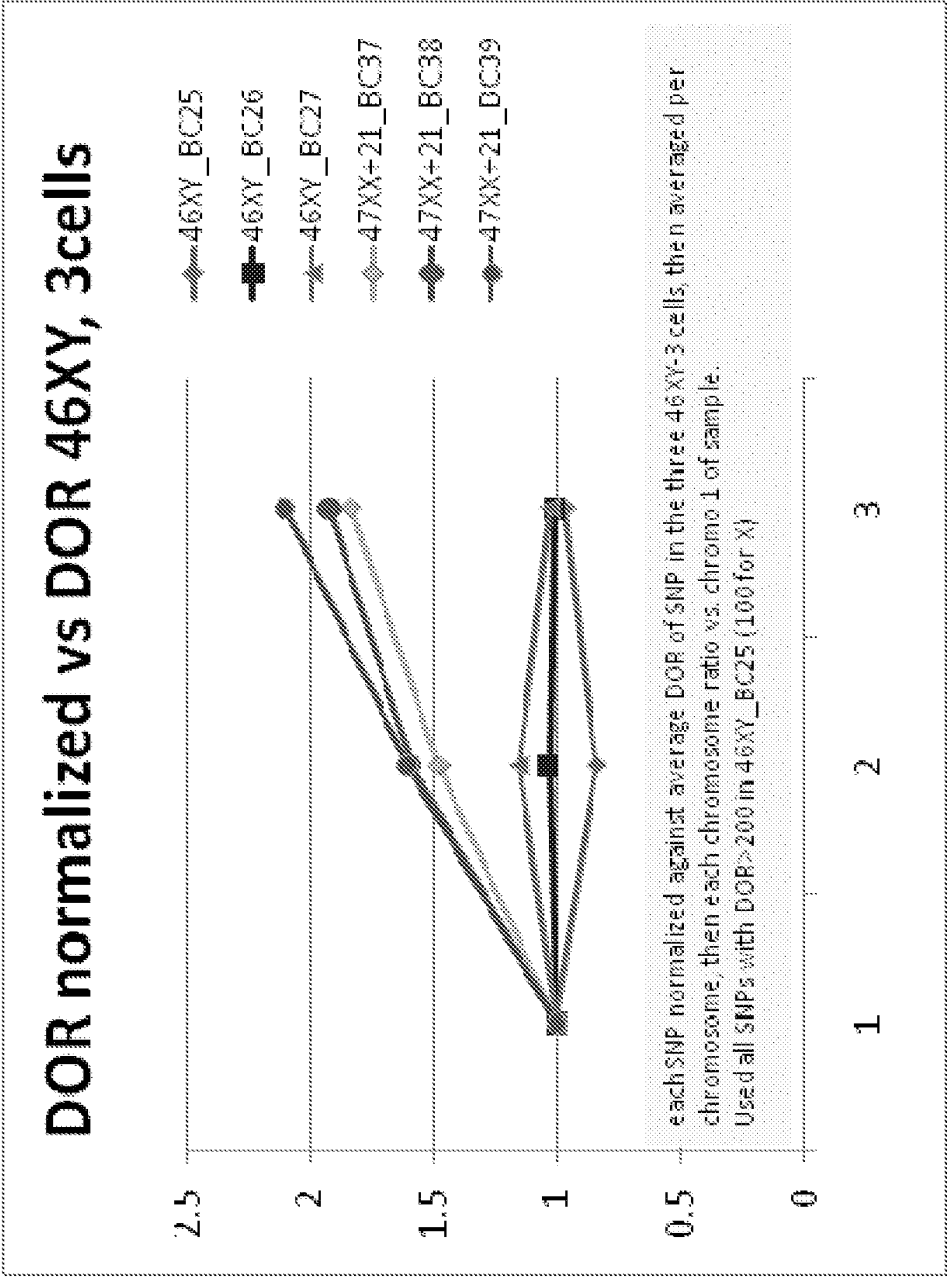


Figure 24

- Data from 3 cells with Arcturus: mini-PCR can detect trisomy 21 from 3 cells.
- Left: chromosome 1, middle chromosome 21 and right chromosome X.
- Sorted from left to right by DOR for each chromosome.

NeoGenomics
4 Page 1351

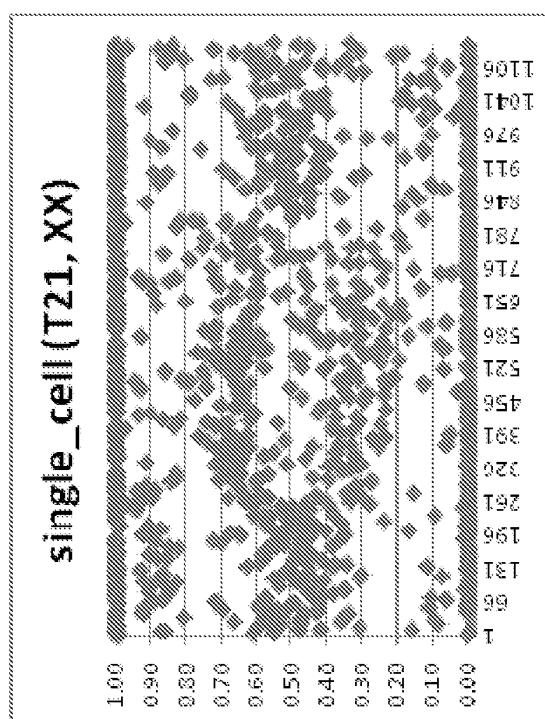
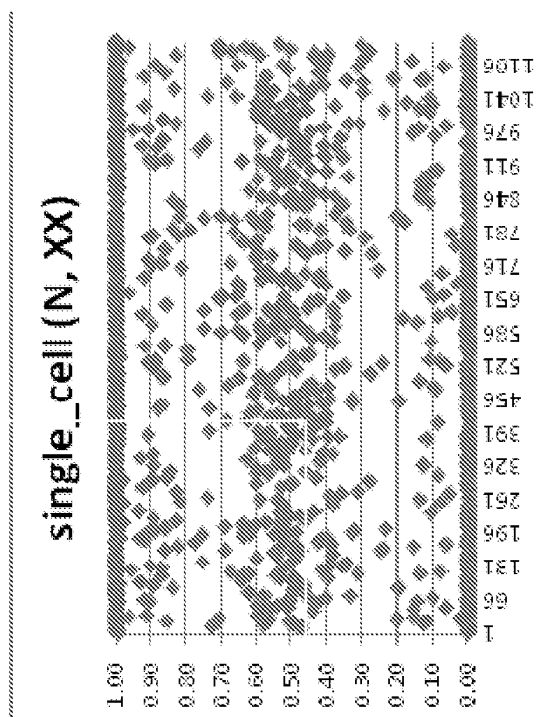


Figure 26

Figure 27
 Minor Allele Frequencies for NIPD Targets, Chromosome 21

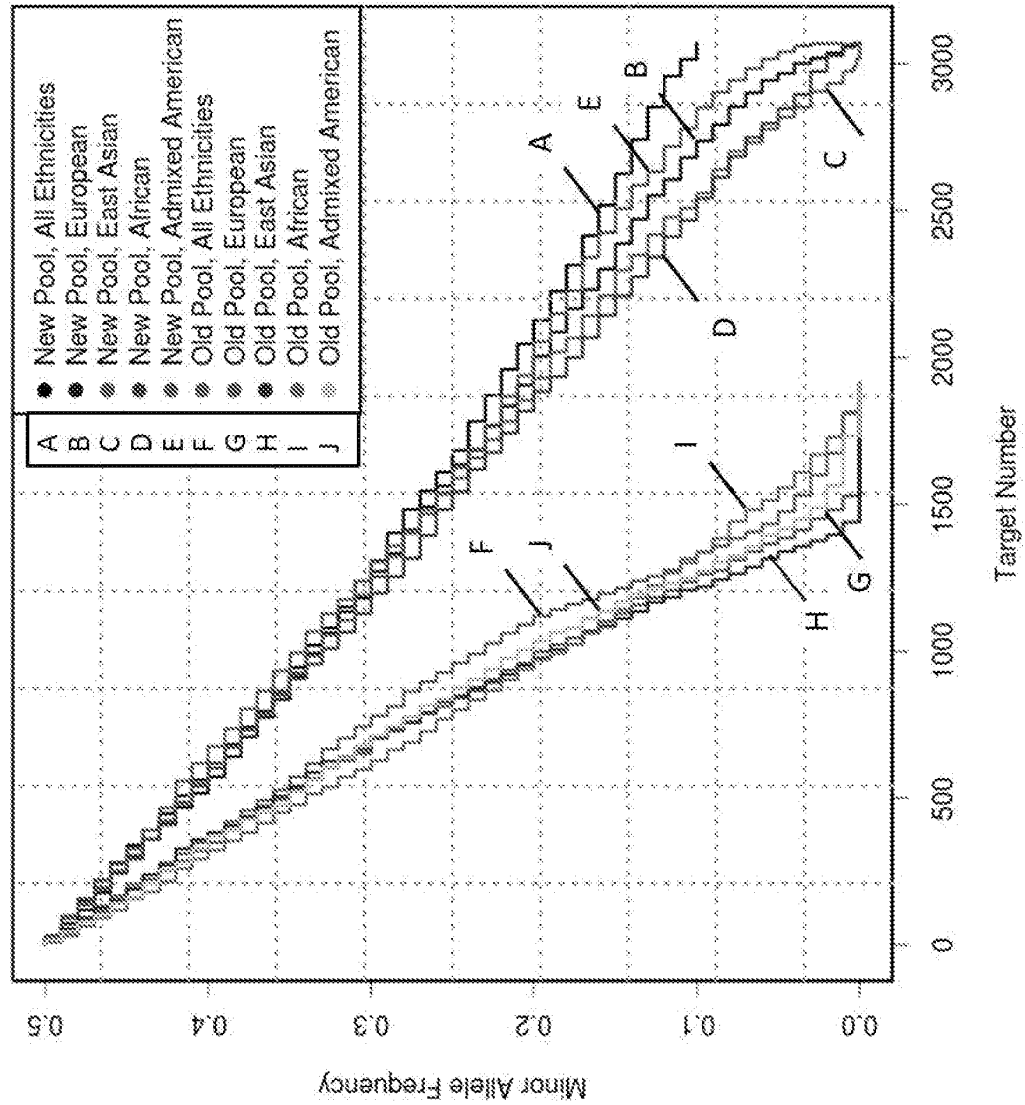
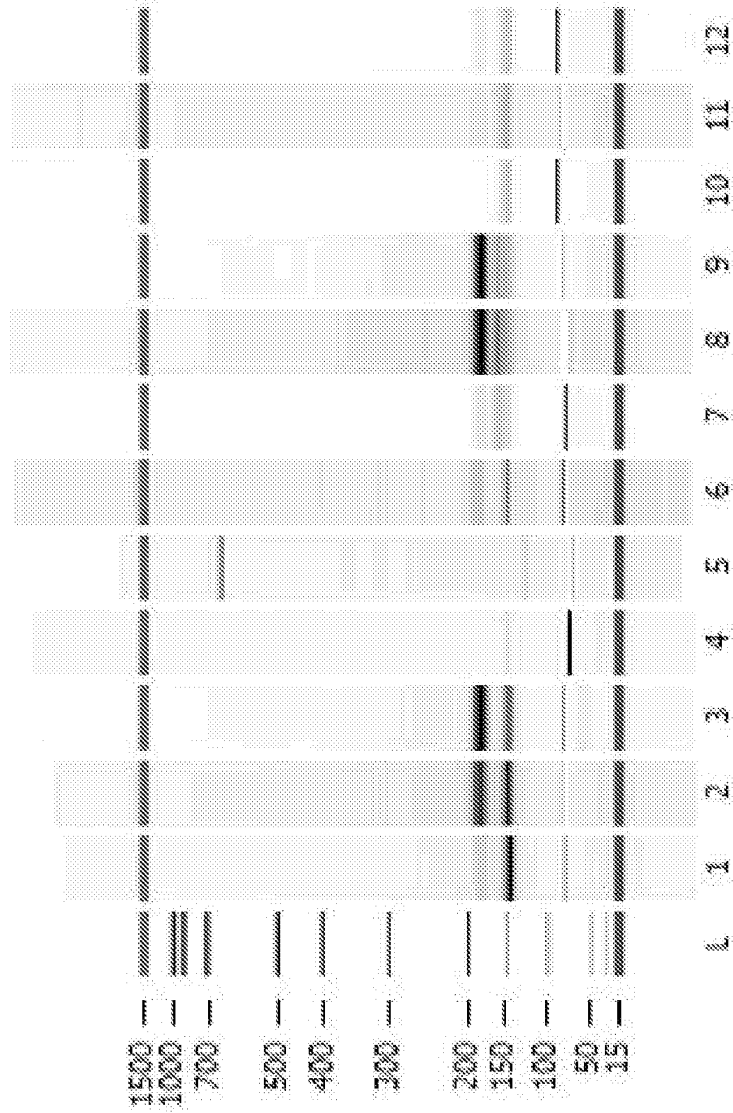


Figure 28A



Direct PCR 2.4k 24ng 10nM

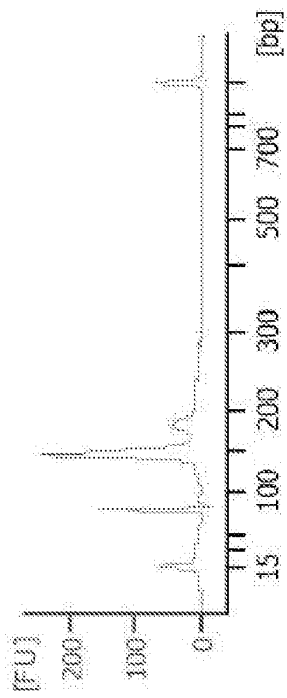


Figure 28B

Direct PCR 2.4k 80ng 10nM

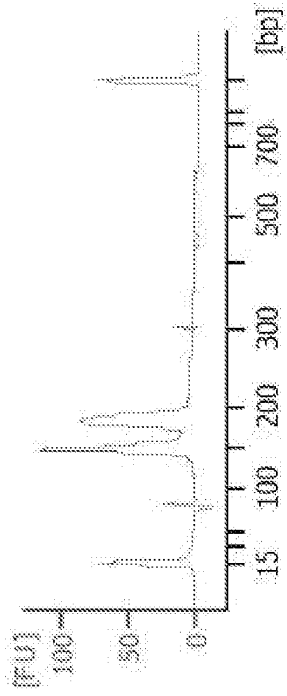


Figure 28C

Direct PCR 2.4k 250ng 10nM

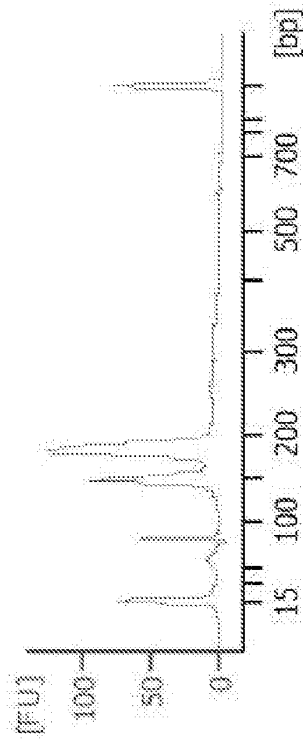


Figure 28D

Direct PCR 2.4k 24ng 1nM

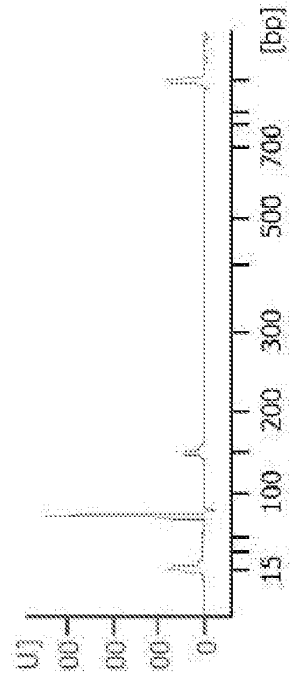


Figure 28E

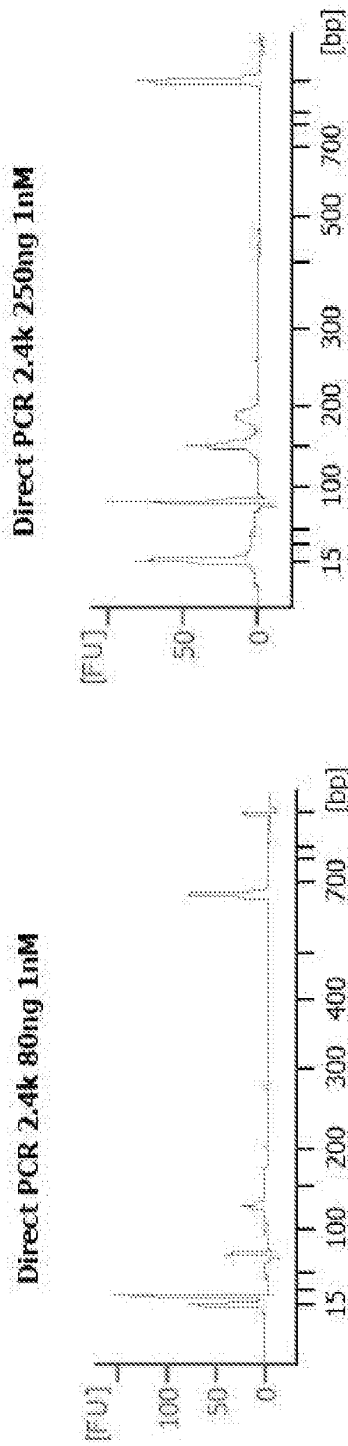


Figure 28F

Figure 28G

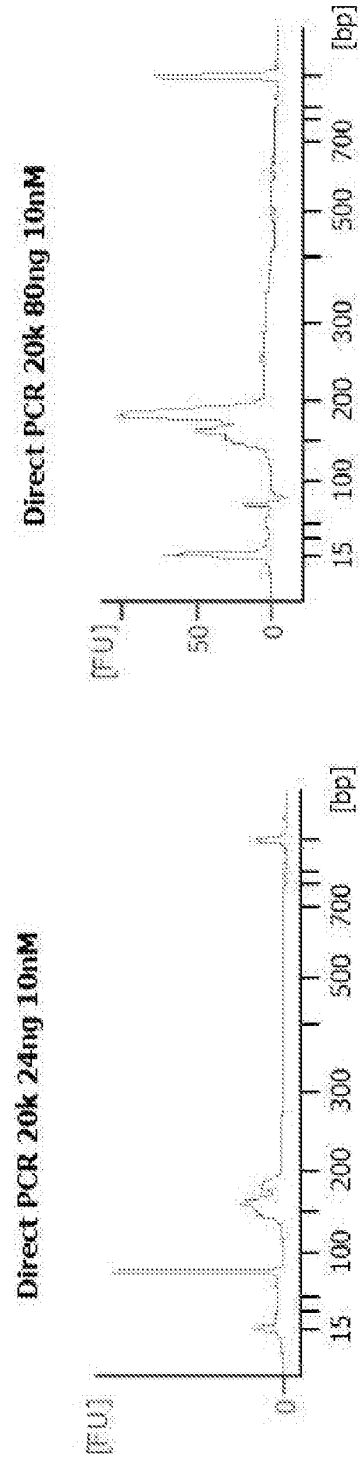


Figure 28H

Figure 28I

Direct PCR 20k 250ng 10nM

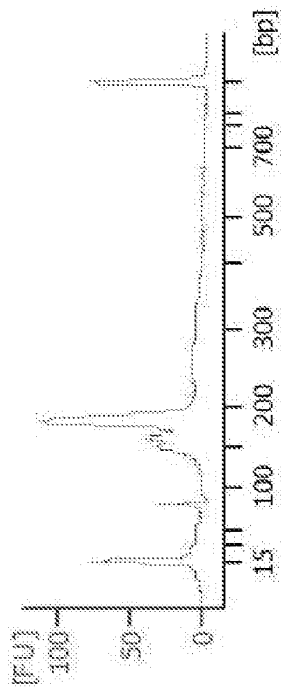


Figure 28J

Direct PCR 20k 24ng 1nM

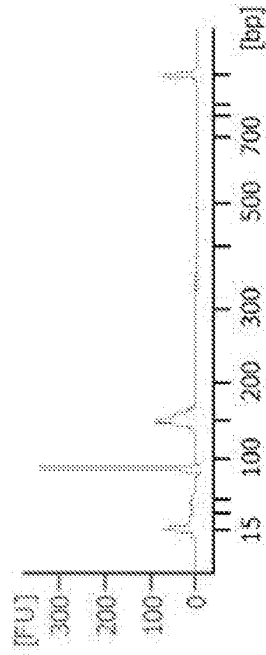


Figure 28K

Direct PCR 20k 80ng 1nM

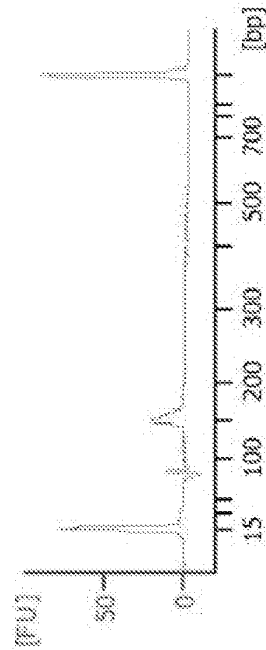


Figure 28L

Direct PCR 20k 250ng 1nM

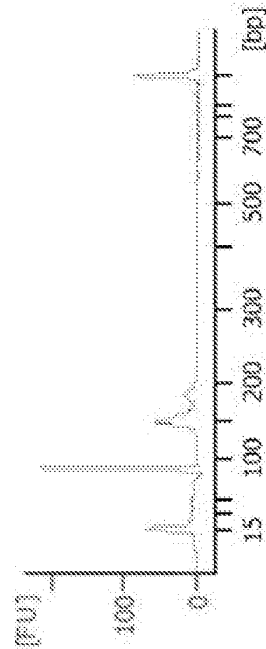
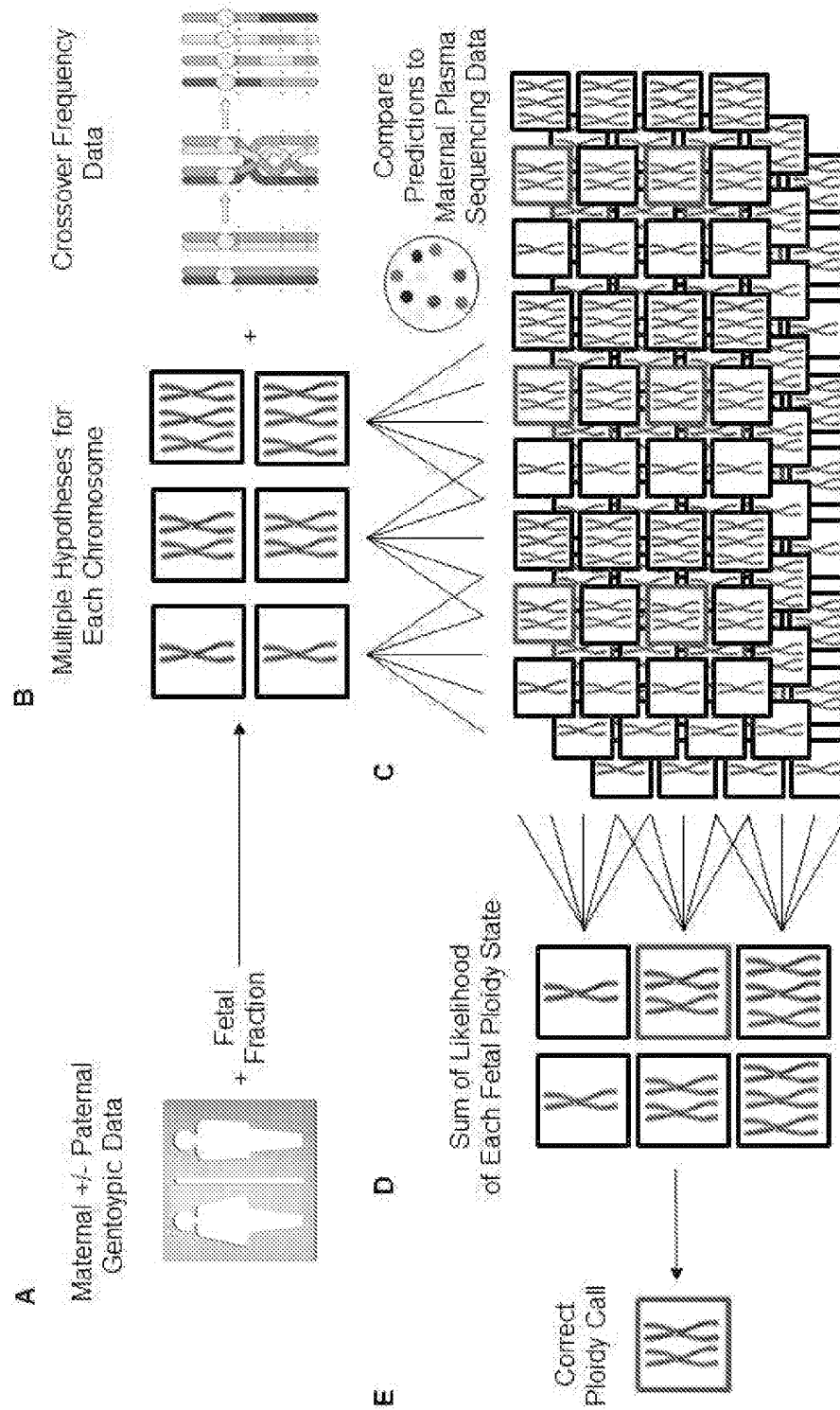
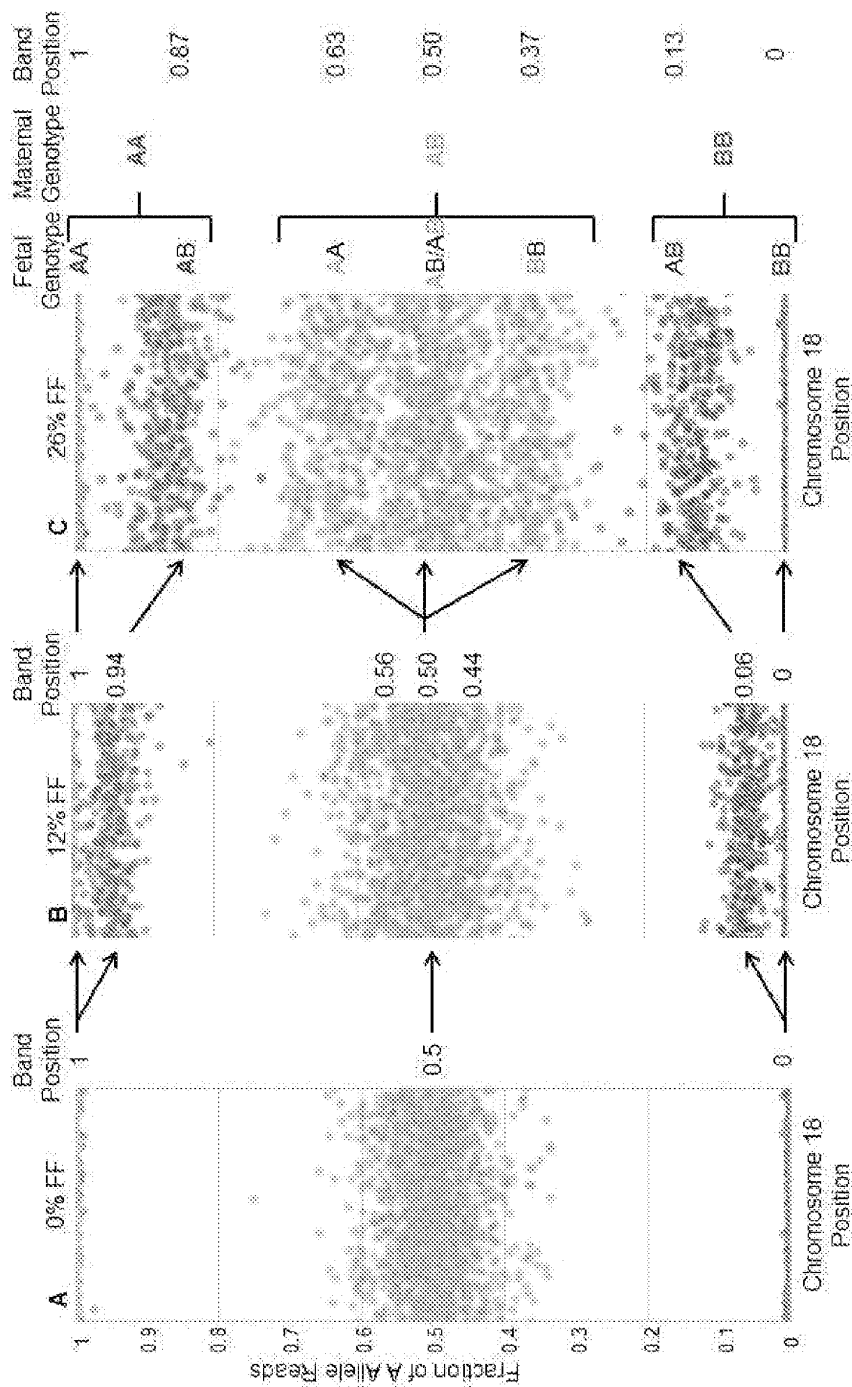


Figure 28M

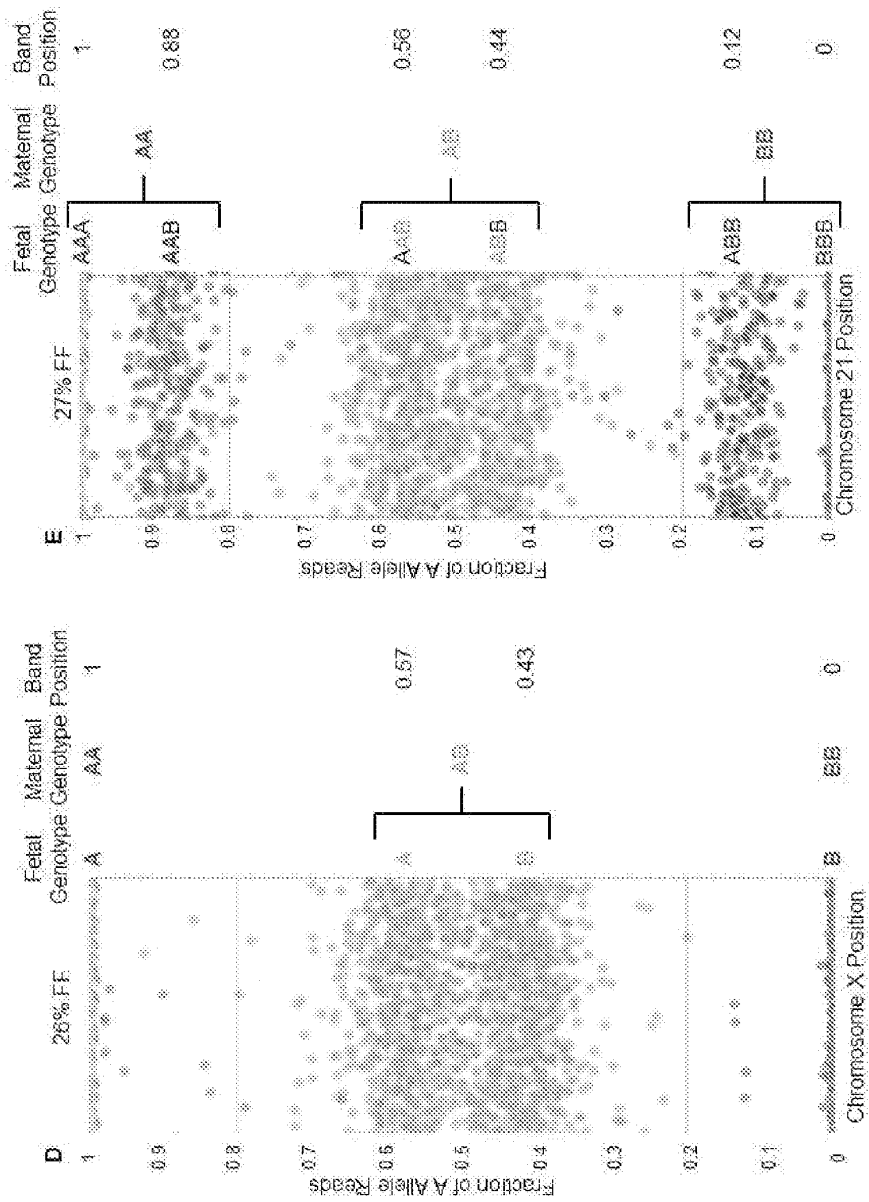
Figures 29A-29E



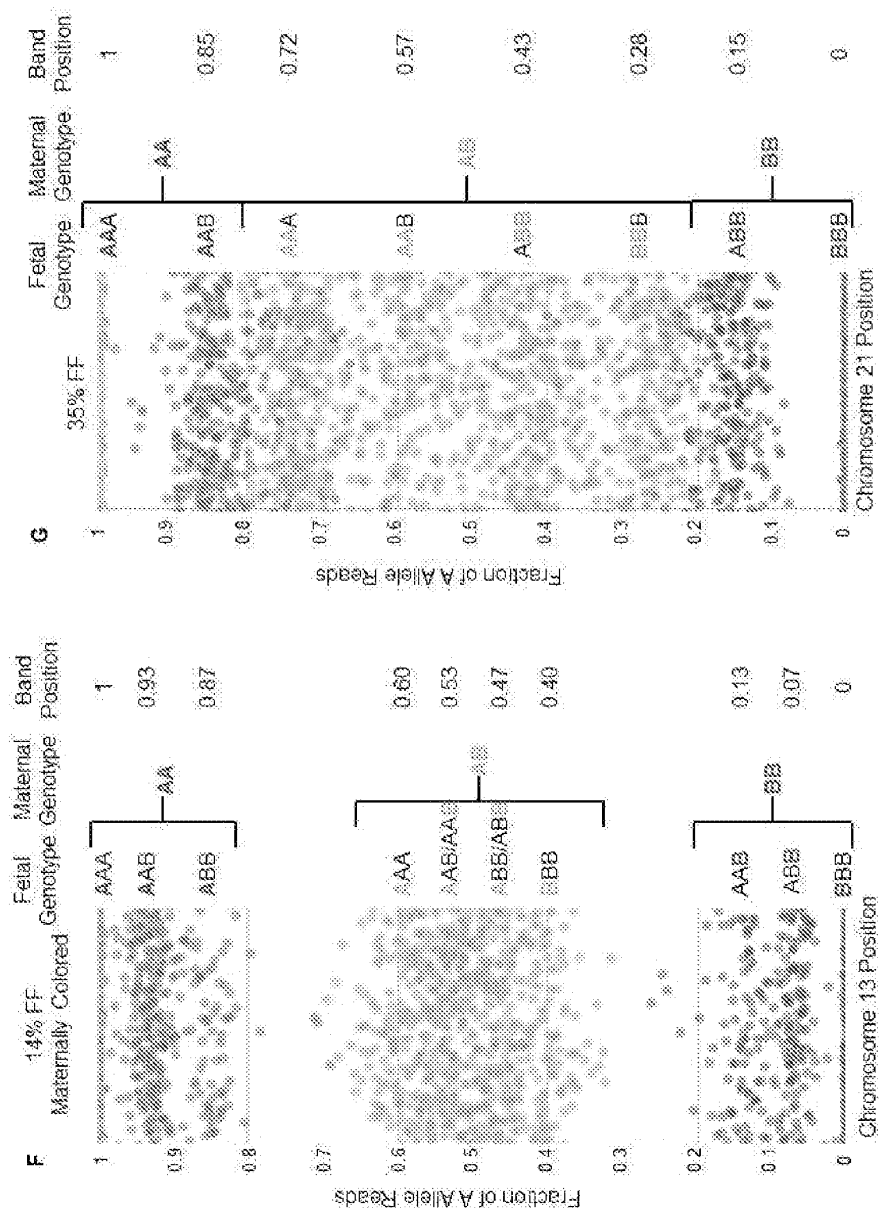
Figures 30A-30C



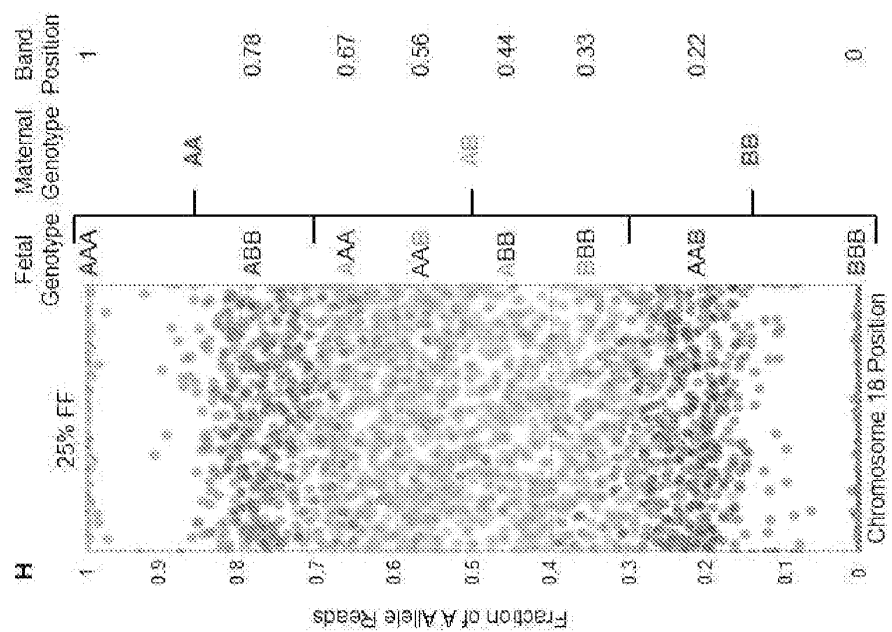
Figures 30D and 30E



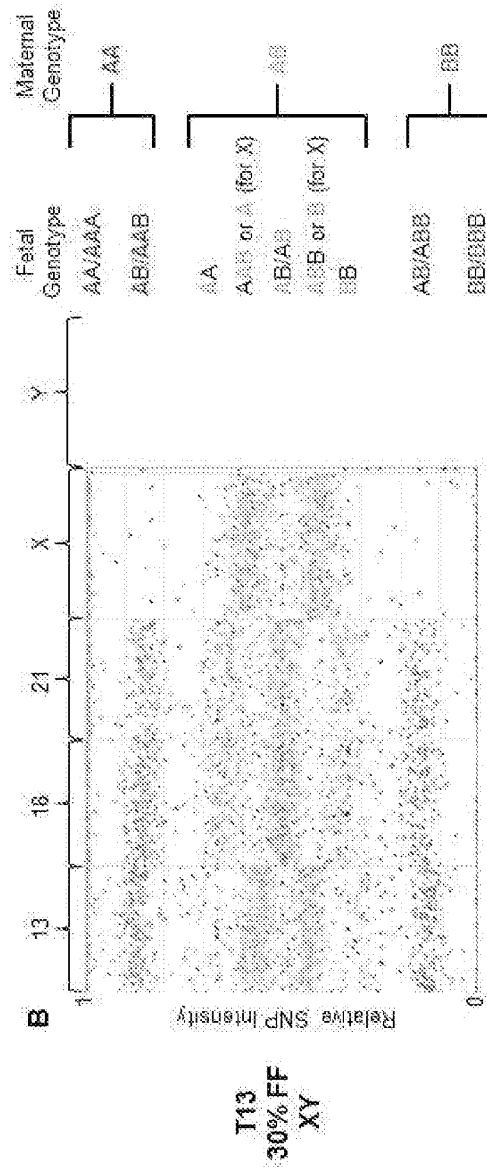
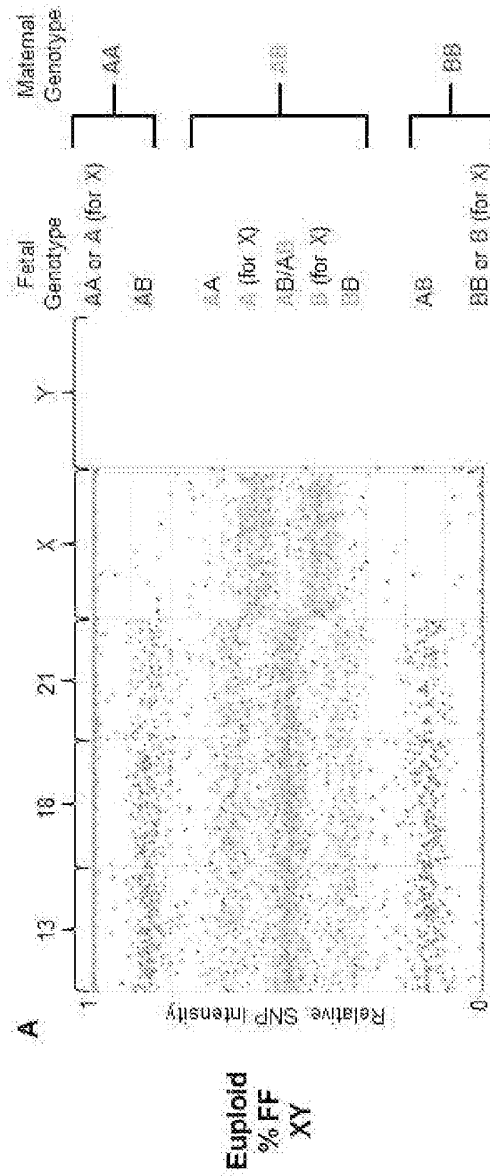
Figures 30F and 30G



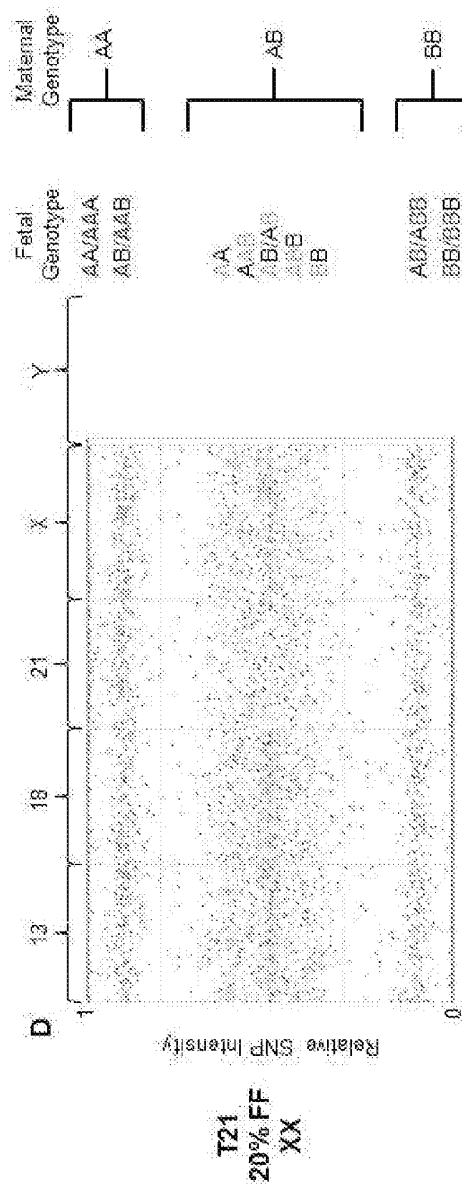
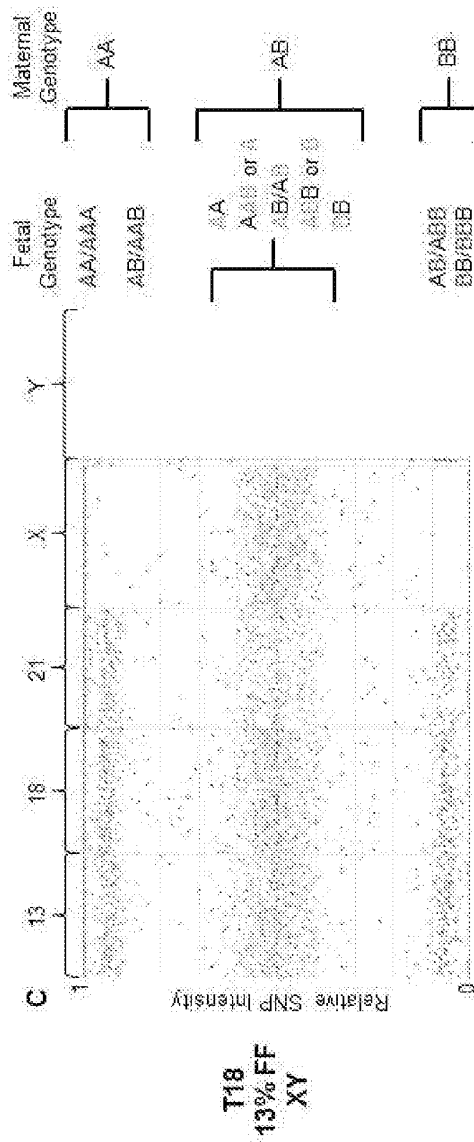
Figures 30H



Figures 31A and 31B



Figures 31C and 31D



Figures 31E and 31F

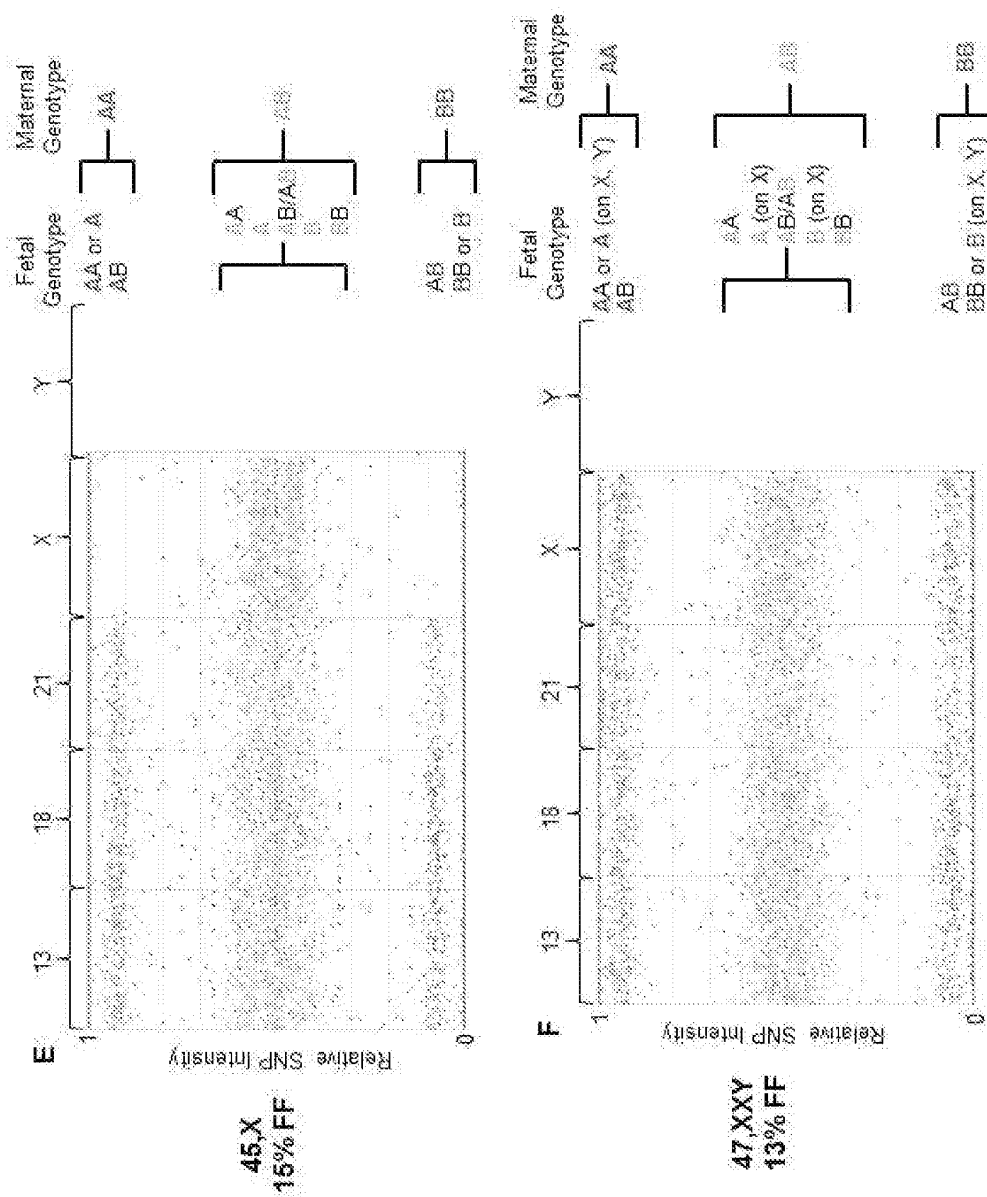


Figure 31G

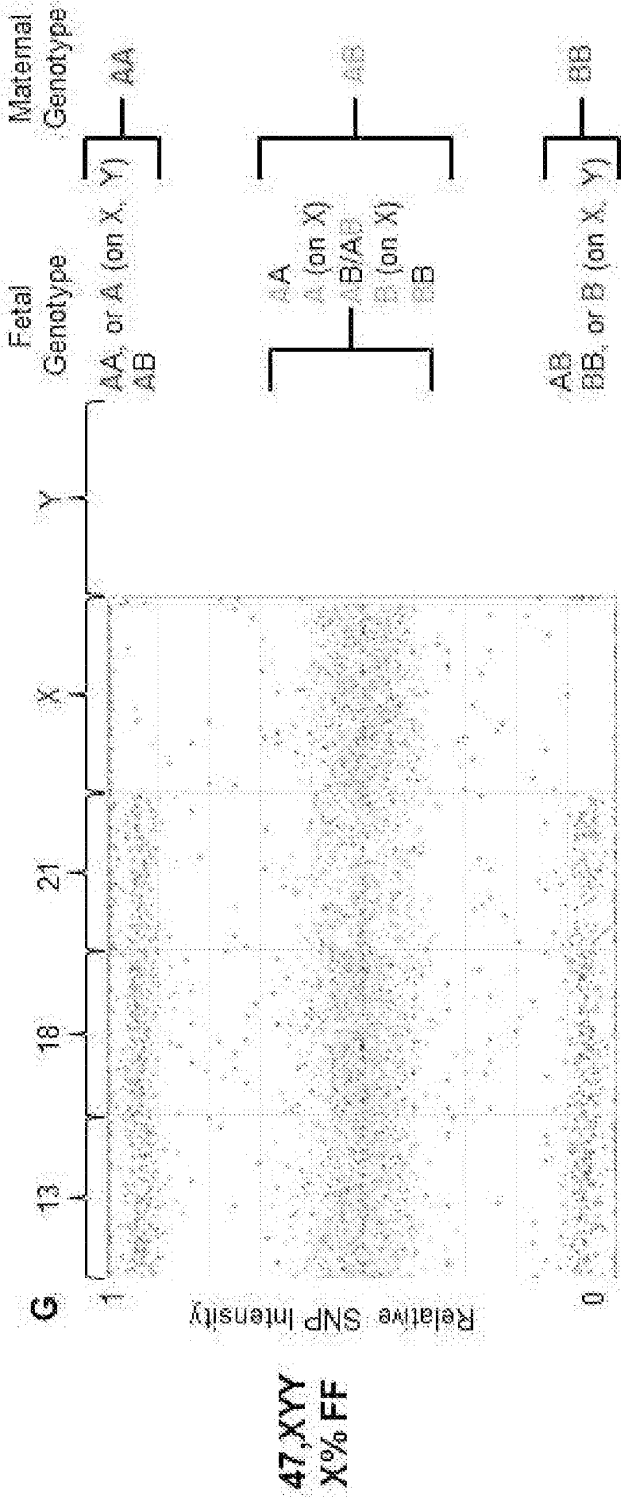
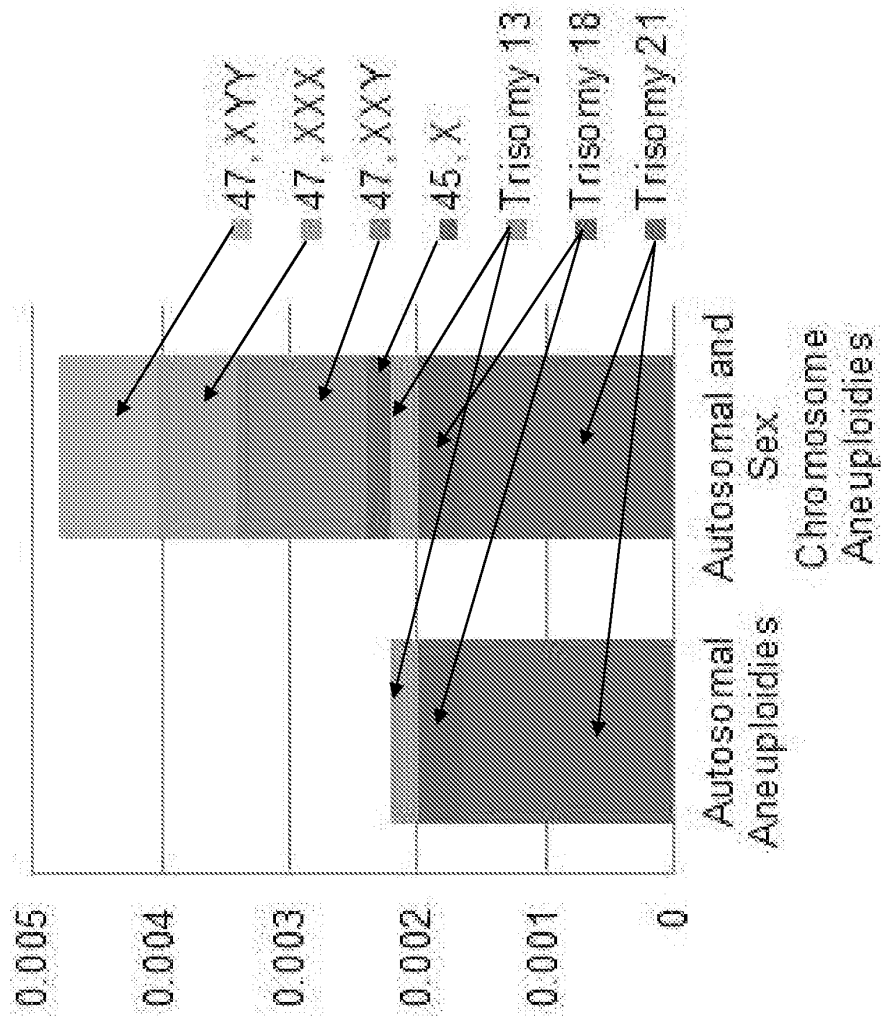


Figure 32



Electronic Patent Application Fee Transmittal				
Application Number:				
Filing Date:				
Title of Invention:		HIGHLY MULTIPLEX PCR METHODS AND COMPOSITIONS		
First Named Inventor/Applicant Name:		Bernhard Zimmermann		
Filer:		Anton J Bokal/suzanne baycan		
Attorney Docket Number:		N.004.US.21		
Filed as Large Entity				
Filing Fees for Utility under 35 USC 111(a)				
Description	Fee Code	Quantity	Amount	Sub-Total in USD(\$)
Basic Filing:				
Utility application filing	1011	1	280	280
Utility Search Fee	1111	1	600	600
Utility Examination Fee	1311	1	720	720
Pages:				
Utility Appl Size fee per 50 sheets >100	1081	3	400	1200
Claims:				
Claims in Excess of 20	1202	8	80	640
Miscellaneous-Filing:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD(\$)
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				
Miscellaneous:				
Total in USD (\$)				3440

Electronic Acknowledgement Receipt	
EFS ID:	23725094
Application Number:	14877925
International Application Number:	
Confirmation Number:	6791
Title of Invention:	HIGHLY MULTIPLEX PCR METHODS AND COMPOSITIONS
First Named Inventor/Applicant Name:	Bernhard Zimmermann
Customer Number:	110711
Filer:	Anton J Bokal/suzanne baycan
Filer Authorized By:	Anton J Bokal
Attorney Docket Number:	N.004.US.21
Receipt Date:	07-OCT-2015
Filing Date:	
Time Stamp:	23:43:03
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	Deposit Account
Payment was successfully received in RAM	\$ 3440
RAM confirmation Number	9667
Deposit Account	506075
Authorized User	BOKAL, ANTON J.
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

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

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Application Data Sheet	ADSaia0014.pdf	1562557	no	10
			81beaa65e66b16c8056b5f644b1bfe263ad6f754		
Warnings:					
Information:					
2		n004us21ApplicationAsFiled.pdf	1121565	yes	239
			cc68267f2f4d1d4241f671a9bad1f8ffa0eddfaa		
	Multipart Description/PDF files in .zip description				
	Document Description		Start	End	
	Specification		1	233	
	Claims		234	238	
	Abstract		239	239	
Warnings:					
Information:					
3	Drawings-only black and white line drawings	DRAWINGS_N004US21.pdf	1984627	no	39
			5b320b85bc73d3c1e75b35cba13b49fae6f6395b		
Warnings:					
Information:					
4	Fee Worksheet (SB06)	fee-info.pdf	38666	no	2
			fc1afd87e07ea31141db291412c6bfed705f80f5		
Warnings:					
Information:					
Total Files Size (in bytes):			4707415		

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

New Applications Under 35 U.S.C. 111

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

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

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

New International Application Filed with the USPTO as a Receiving Office

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