Noninvasive Detection of Fetal Genetic Variations through Polymorphic Sites Sequencing of Maternal Plasma DNA

Song Gao*

The State Key Laboratory Breeding Base of Basic Science of Stomatology & Key Laboratory of Oral Biomedicine Ministry of Education, School & Hospital of Stomatology, Wuhan University, Wuhan, China.

*Corresponding author

Song Gao
Associate Investigator
School & Hospital of Stomatology
Wuhan University
237 Luoyu Road, Wuhan, China, 430079
Email: gaos@whu.edu.cn

Abstract

Non-invasive prenatal testing (NIPT) for common fetal aneuploidies using circulating cell free DNA in maternal plasma has been widely adopted in clinical practice for its sensitivity and accuracy. However, the detection of subchromosomal abnormalities or monogenetic variations using such a method showed no cost-effectiveness or satisfactory accuracy. Here we show that with the aid of polymorphic sites sequencing, fetal fraction of the sample and genotype of the target site were determined with high accuracy. Then genetic variations at the chromosomal, subchromosomal and nucleotide levels were detected using the overall allelic goodness-of-fit test of all target polymorphic sites to each possible genetic model. Finally, relative allelic distributions for each amplicon were visualized and genetic variations at the chromosomal, subchromosomal and nucleotide levels were determined by distinct characteristic clusters on allelic distribution plot of each possible genetic model. As no parental genetic information was required and all allelic information retained for amplicon sequencing, the reported approach has the potential to simultaneously detect genetic variations at different levels, facilitating the extension of NIPT to all common genetic conditions for general low-risk pregnancies and target variations for certain high-risk pregnancy groups.

Keywords: NIPT, amplicon sequencing, goodness of fit, polymorphic site, noninvasive prenatal testing, fetal fraction

Introduction

Noninvasive prenatal testing (NIPT) is now widely used for the detection of fetal chromosomal aneuploidies and certain copy number variations, where cell-free DNA (cfDNA) in maternal plasma[1] was analyzed by whole-genome sequencing (WGS) [2,3], single-nucleotide polymorphism (SNP)[4,5] or microarray[6,7]. NIPT showed high test sensitivity and specificity for common fetal aneuploidies, such as trisomies 21, 18 and 13, but low in detecting subchromosomal deletions and duplications[8-10], especially when the genomic aberrations were small[11-15]. For monogenic disorders, different noninvasive approaches have been developed[16], but the application of such methods in clinical practice has lagged behind aneuploidy testing due to high costs and technical challenges.

In cfDNA, a certain number of polymorphic sites showed allelic imbalance due to the presence of fetal DNA. When the fetus inherits a paternal allele different from the mother's (Fig. S1), fetal aneuploidies can be detected using relative allelic counts. As maternal-fetal genotype information for each polymorphic site is encoded in the imbalanced allelic counts, genetic variations could be determined reliably by analyzing allelic imbalances of groups of polymorphic sites, potentially simultaneous detection of chromosome aneuploidies, subchromosomal deletions and duplications, or nucleotide disorders in a low-cost and high-accuracy manner.

Results

Fetal Fraction Estimation

A panel of insertion/deletion polymorphic markers was PCR amplified using maternal cfDNA as template[17], and the possible maternal-fetal genotype for each marker was estimated using its allelic read counts (Fig. S2). Then, reads counts of fetal origin were estimated for each amplicon and fetal fraction calculated for each sample by fitting a robust linear regression model (Table S1, Fig. S3 and Supplementary Information). A high degree of correlation was observed for fetal fractions estimated this way and that estimated using WGS sequencing when samples with low mapped bin counts were excluded (Fig. S3). Similarly, nearly identical fetal fraction estimates were observed for library- or sequencing-level replicates[18] (Fig. 1), indicating that the method for fetal fraction estimation was accurate and reliable, although estimation accuracies were affected by both the abundance of fetal materials and the sequencing coverage (Fig. S4).

Genotype Estimation for Polymorphic Site

When both the mother and the fetus are normal diploid, one of five maternal-fetal genotypes is possible for each polymorphic site in maternal plasma DNA, and the genotype could be estimated by allelic goodness-of-fit test[19] where the observed allelic counts were tested against the corresponding expected allelic counts for each possible genotype model (Fig. S5). As the raw ΔAIC used for model selection was highly influenced by both fetal fractions and total allelic counts (Fig. S6), adjusted ΔAIC was calculated and nearly similar magnitude of adjusted values observed even for polymorphic sites with different fetal fractions and total allelic counts, indicating that the adjusted ΔAIC could be a good measure when checking the fitness of different genotype models to polymorphic sites with different sequencing depths. In addition, when

estimating maternal-fetal genotype of each polymorphic site with allelic goodness-of-fit test, the true underlying genotype model should be included in the analysis and prior knowledge of the target polymorphic site was desired. As expected, more than 95% of the maternal-fetal genotypes could be correctly estimated for polymorphic sites of simulated cfDNA samples when the sequencing coverage was ≥2000 and the fetal fraction was ≥0.05 (Fig. S7).

Detection of Chromosomal Aneuploidies

When there was a fetal aneuploidy, all polymorphic sites on the target chromosome were affected. Therefore for each polymorphic site, a best fit genotype was calculated for each possible maternal-fetal aneuploidy model, and the model that showed overall best fit to all polymorphic sites on the target chromosome was chosen followed with fetal aneuploidy determination accordingly (Fig. S5). Such an approach might seem unsound mathematically, but were sensitive and reliable to detect chromosomal aneuploidies for our simulated samples, possibly due to its similarity to repeated tests of goodness-of-fit[19] where each polymorphic site was considered as an experimental repetition. As the majority of target polymorphic sites were informative for estimating fetal aneuploidies (Fig. S8-9) except for one-allele sites, all normal and aneuploidy chromosomes were correctly identified (Fig. S10, Fig. 2a-b) when samples with both low sequencing coverage and low fetal fraction were excluded. When detecting fetal aneuploidies, all possible maternal-fetal aneuploidies for the target chromosome should be checked, as is the case for detecting sex chromosome aneuploidies, where both the normal (XX and XY)

and all five of the better-known sex aneuploidies (XO, XXY, XXX, XYY and XXYY)[20] should be considered.

Detecting Subchromosomal Abnormalities

As the heterozygotes for some subchromosomal microdeletions[21] or microduplications[22] could be phenotypically normal, heterozygous or homozygous subchromosomal abnormality models for either the mother, the fetus or both should be tested if necessary. Based on statistics, some target region could be tested using models assuming that the mother was homozygous normal for the subchromosomal abnormality as if checking fetal aneuploidy, while other target region should be tested using models assuming that the mother was heterozygous or homozygous for the disease (Fig. S11-12). As one-allele polymorphic sites were not informative for detecting target microdeletions, the overall fit for monosomy-nullisomy model could not be distinguished from that of the nullisomy-monosomy model (Fig. S11). Therefore, alternative noninvasive or invasive approaches for accurately detecting such disease conditions should be performed when necessary. As expected, subchromosomal microdeletions or microduplications could be detected with accuracy when at least two alleles were detected for some polymorphic sites in the target region (Fig. 2, Fig. S11-12). When best overall fits to both a disomic-disomic model and a tetrasomic-tetrosomic model were observed for a microduplication, the target was estimated to be disomic-disomic, as any genotype in the disomy-disomy model had a corresponding counterpart in the tetrasomy-tetrasomy model with identical relative allelic distributions (for example, AB|AA corresponds to AABB|AAAA). However, if only a

tetrasomic-tetrasomic model showed overall best fit for all polymorphic sites on the target microduplication region, tetrasomy-tetrasomy was estimated (Fig. S12).

Detection of Short Genetic Variations

Single-base-pair substitutions, small (≤20bp) deletions, small (≤20bp) insertions and small (≤20bp) indels are the major types of mutations associated with human inherited diseases reported in the Human Gene Mutation Database (HGMD)[23]. To detect such genetic variations in cfDNA samples, each target site was amplified and its genotype estimated using allelic goodness-of-fit test (Fig. 3a). Then the nucleotide sequence of each target allele was checked and the wildtype-mutant genotype of each target site was determined accordingly (Table S2-3). As the accuracy for genotype estimation using a single allelic site was not perfect (Fig. S7), library or sequencing level repeats were desired as demonstrated by the replication dataset (Fig. S13). On the other hand, sequencing with replicates increased the overall cost considerably, and it was not cost effective when used for detecting genetic mutations with low disease prevalence. Therefore, a limited number of replicates were suggested initially for each target site with low disease prevalence, and if mutant alleles were detected for a target site, further analysis and possibly retesting using more replicates were performed. Such a two-tier test strategy could reduce the overall cost greatly and increase the positive predictive value (PPV), as only a small number of target sites were to be retested and disease incidences were increased for retested targets.

Graphical Analysis of Genetic Variations

In each cfDNA sample, fetal fractions for all polymorphic sites were considered the same even though their chromosomal positions were different (Table S2-3). As only a limited number of genotypes were possible, and the relative allelic abundances for each genotype were determined by the sample's fetal fraction, distinct clusters were observed when the second most abundant relative allelic count was plotted against the most abundant relative count for each polymorphic site (Fig. S14). In addition, the characteristic cluster distribution for all polymorphic sites in a sample was informative enough to identify genetic abnormalities at the chromosomal or subchromosomal level (Fig. S15-18 and Fig. 3). To detect nucleotide-level genetic variations for a single target site, both the wild-type and the mutant alleles were counted first, followed with the plotting of the most abundant relative mutant allelic count against the relative count of the wildtype. Subsequently, maternal-fetal genotype of the target site was estimated by eye examination of its characteristic allelic distribution (Fig. S19, Table S2-3 and Fig. 3b). When allelic clusters were not in the expected positions, either the true model was excluded and wrong model was fitted or there were non-random outliers. In such a case, further analyses, optimizing test routines or checking additional models should be followed.

Discussion

Currently, cfDNA based NIPT approaches have been widely available for detecting fetal aneuploidies[2-5], subchromosomal abnormalities[12,15] or monogenic diseases[24-30] in clinical practice. However, no approach reported could detect genetic variations simultaneously at both the chromosomal/subchromosomal level and the nucleotide level. Here we reported the simultaneous detection of genetic abnormities at different levels by amplicon

sequencing of polymorphic sites and specific targets. As nearly all genetic abnormalities were identified correctly, high sensitivity and specificity was observed for our simulated samples. Although different sensitivities were reported when detecting genetic abnormalities at different levels[31], the sensitivity for our reported approach should not varied much, as all relative allelic information were encoded in amplified amplicons for each polymorphic site and different alleles of each amplicon had nearly identical sequences with similar amplification properties.

In clinical settings, accuracy, specificity and sensitivity should be addressed using real cfDNA samples, as clinical data was inherently noisy and discrepancy between the genotypes of maternal plasma fetal DNA and the fetal genome were reported in some samples possibly due to confined placental mosaicism[32]. Moreover, prior knowledge about the detecting targets was required and no off-target variations could be detected[33], while WGS-based NIPT methods could detect incidental variations with no additional cost.

In principle, target amplicon sequencing could be applied to detect other genetic variations as well. For examples, chromosomal inversion or translocation with known break point could be detected by amplicons covering the specific breakpoint. Genomic abnormalities for preimplantation embryos or non-pregnant samples could be detected using polymorphic sites sequencing as well, as distinct allelic distributions for all target polymorphic sites were informative enough to identify different abnormalities (Fig. S20). For cfDNA sample from a surrogate mother, fetal fraction was estimated first using a panel of polymorphic sites and goodness-of-fit test aided by iteratively updated estimates

(Fig. S21), then genetic variations could be detected by checking all possible genotype models. For samples from a mother with multiple pregnancies, fetal fraction for each fetus could be estimated using a similar approach (Fig. S21), where each fetal fraction estimate was updated iteratively until converge, and genetic abnormalities could be detected using allelic goodness-of-fit test as expected allelic counts for each polymorphic site could be calculated when fetal fractions for all fetuses were available.

Collectively, nearly all common genetic disorders could be detected with the aid of amplicon sequencing, and expansion of NIPT to detect both genetic conditions that were common to all pregnancies and disorders that had high prevalence in particular groups would have great socioeconomic benefits.

Methods

Dataset

The insertion/deletion polymorphism[17] dataset (BioProject ID: PRJNA387652) and the replication[18] dataset (BioProject ID: PRJNA517742) were retrieved from the NCBI SRA database. The simulated datasets were generated using ART[34] simulator (see Supplementary Information for detailed descriptions).

Reads Processing and Mapping

Reads retrieved from SRA or simulated were filtered out using custom scripts where the low quality bases were removed and the longest subsequence in each read was retained so that all bases had a quality score greater than 14. Whole genome sequencing reads were

mapped by bowtie2[35]. For amplicon reads, one or several unique 12-mer indexes were extracted from each amplicon and each read was mapped to an amplicon using such indexes. Then the allelic reads for each amplicon in each sample were counted using unique allelic sequences.

Fetal Fraction Estimation by Allelic Read Counts

For amplicon sequencing data, fetal fractions were estimated as follows. For each polymorphic site, read counts for all alleles were sorted in descending order and labeled as R1, R2, R3, etc. Then the possible maternal-fetal genotype was estimated using allelic read counts (Fig. S2) followed by the estimation of fetal and total read counts (Table S1). Finally, fetal fraction was estimated using fetal and total read counts and a robust linear regression model (see Supplementary Information).

Fetal Fraction Estimation by Whole Genome Sequencing

Fetal fraction was calculated as described using the formula[36]

Fetal Fraction $(f)=\frac{2.0\times \mathrm{med}(\mathrm{Chr_Y})}{\mathrm{med}(\mathrm{Chr_X})+\mathrm{med}(\mathrm{Chr_Y})}$, where $\mathrm{med}(\mathrm{Chr_X})$ and $\mathrm{med}(\mathrm{Chr_Y})$ represent the median read counts of the 50-kb bins on the X and Y chromosomes, respectively. Briefly, the 50-kb bins from the X and Y chromosomes were extracted and bins having too low or too high read counts were filtered out using 200 whole genome sequencing samples (SRR6040419-SRR6040618) from the project PRJNA400134[37] as follows. Reads were firstly mapped to the human reference genome using bowtie2[35], and total reads mapped into each X or Y bin were counted for each sample using custom scripts. Subsequently, X bins containing no mapped reads in more than 25% of the samples or containing read counts not in the range of $\mathrm{Median}\pm 3.0\times \mathrm{MAD_e}$ were removed, while Y bins containing at

least one read in more than 25% of the female pregnancies, containing no mapped reads in more than 25% of the male pregnancies or containing read counts not in the range of Median $\pm 3.0 \times MAD_e$ were removed as well. Hence, a total of 2760 chromosome X bins and 192 chromosome Y bins were identified as informative bins. Fetal fractions for the 61 samples from PRJNA387652 were calculated using the median count values of X bins and Y bins as described above.

Maternal-Fetal Genotype Estimation

Fetal fraction was estimated first for each sample. Then for each polymorphic site, reads for each allele were counted, followed by the calculations of AICs for all possible genotype models using goodness-of-fit test. Finally, the genotype for each polymorphic site was estimated to be the one with the minimal AIC, and ΔAIC was calculated as the absolute difference between the minimal AIC and the second minimal AIC. To detect chromosomal or subchromosomal genotypes, the minimal AICs for all allelic sites was averaged for each chromosomal/subchromosomal model, and the chromosomal/subchromosomal genotype was estimated to be the one associated with the minimal average AIC (see Supplementary Information for detailed descriptions).

Statistical Analysis

Statistical analysis was performed in R[38] (version 3.5.1). AICs were calculated using custom scripts.

Data availability

The author declares that the data supporting the findings of this study are available within the paper and its supplementary information file. Source data for all figures and scripts for analyzing the data are available from the author upon request.

Acknowledgements

The author thanks Yongxin Ke for technical and administrative assistance, discussions and comments on the project.

Author contributions

S. G. designed the experiments, performed the experiments, analyzed the data and wrote the manuscript.

Competing interests

A patent application has been filed relating to this project.

Supplementary Information is available for this paper.

References

1 Lo, Y. M. *et al.* Presence of fetal DNA in maternal plasma and serum. *Lancet* **350**, 485-487, doi:10.1016/s0140-6736(97)02174-0 (1997).

2 Fan, H. C., Blumenfeld, Y. J., Chitkara, U., Hudgins, L. & Quake, S. R. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 16266-16271, doi:10.1073/pnas.0808319105 (2008).

3 Chiu, R. W. *et al.* Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 20458-20463, doi:10.1073/pnas.0810641105 (2008).

4 Zimmermann, B. *et al.* Noninvasive prenatal aneuploidy testing of chromosomes 13, 18, 21, X, and Y, using targeted sequencing of polymorphic loci. *Prenat Diagn* **32**, 1233-1241, doi:10.1002/pd.3993 (2012).

5 Liao, G. J. *et al.* Noninvasive prenatal diagnosis of fetal trisomy 21 by allelic ratio analysis using targeted massively parallel sequencing of maternal plasma DNA. *PLoS One* **7**, e38154, doi:10.1371/journal.pone.0038154 (2012).

6 Srebniak, M. I. *et al.* The influence of SNP-based chromosomal microarray and NIPT on the diagnostic yield in 10,000 fetuses with and without fetal ultrasound anomalies. *Hum Mutat* **38**, 880-888, doi:10.1002/humu.23232 (2017).

7 Juneau, K. *et al.* Microarray-based cell-free DNA analysis improves noninvasive prenatal testing. *Fetal Diagn Ther* **36**, 282-286, doi:10.1159/000367626 (2014).

8 Hu, H. *et al.* Noninvasive prenatal testing for chromosome aneuploidies and subchromosomal microdeletions/microduplications in a cohort of 8141 single pregnancies. *Human Genomics* **13**, 14, doi:10.1186/s40246-019-0198-2 (2019).

9 Srebniak, M. I. *et al.* Social and medical need for whole genome high resolution NIPT. *Mol Genet Genomic Med* **8**, e1062, doi:10.1002/mgg3.1062 (2020).

10 Advani, H. V., Barrett, A. N., Evans, M. I. & Choolani, M. Challenges in non-invasive prenatal screening for sub-chromosomal copy number variations using cell-free DNA. *Prenat Diagn* **37**, 1067-1075, doi:10.1002/pd.5161 (2017).

11 Chau, M. H. K. *et al.* Characteristics and mode of inheritance of pathogenic copy number variants in prenatal diagnosis. *Am J Obstet Gynecol*, doi:10.1016/j.ajog.2019.06.007 (2019).

12 Lo, K. K. *et al.* Limited Clinical Utility of Non-invasive Prenatal Testing for Subchromosomal Abnormalities. *Am J Hum Genet* **98**, 34-44, doi:10.1016/j.ajhg.2015.11.016 (2016).

13 Yu, S. C. *et al.* Noninvasive prenatal molecular karyotyping from maternal plasma. *PLoS One* **8**, e60968, doi:10.1371/journal.pone.0060968 (2013).

14 Li, R. *et al.* Detection of fetal copy number variants by non-invasive prenatal testing for common aneuploidies. *Ultrasound Obstet Gynecol* **47**, 53-57, doi:10.1002/uog.14911 (2016).

15 Yin, A. H. *et al.* Noninvasive detection of fetal subchromosomal abnormalities by semiconductor sequencing of maternal plasma DNA. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 14670-14675, doi:10.1073/pnas.1518151112 (2015).

16 Allen, S., Young, E. & Gerrish, A. in *Noninvasive Prenatal Testing (NIPT)* (eds Lieve Page-Christiaens & Hanns-Georg Klein) 157-177 (Academic Press, 2018).

17 Barrett, A. N. *et al.* Measurement of fetal fraction in cell-free DNA from maternal plasma using a panel of insertion/deletion polymorphisms. *PLoS One* **12**, e0186771, doi:10.1371/journal.pone.0186771 (2017).

18 Kim, J. *et al.* The use of technical replication for detection of low-level somatic mutations in next-generation sequencing. *Nat Commun* **10**, 1047, doi:10.1038/s41467-019-09026-y (2019).

19 McDonald, J. H. *Handbook of Biological Statistics*. Third edn, (Sparky House Publishing, Baltimore, Maryland, 2014).

20 Skuse, D., Printzlau, F. & Wolstencroft, J. Sex chromosome aneuploidies. *Handb Clin Neurol* **147**, 355-376, doi:10.1016/b978-0-444-63233-3.00024-5 (2018).

21 Milili, M. *et al.* A new case of autosomal recessive agammaglobulinaemia with impaired pre-B cell differentiation due to a large deletion of the IGH locus. *European Journal of Pediatrics* **161**, 479-484, doi:10.1007/s00431-002-0994-9 (2002).

22 Ceylan, A. C. *et al.* Autosomal recessive spinocerebellar ataxia 18 caused by homozygous exon 14 duplication in GRID2 and review of the literature. *Acta Neurol Belg*, doi:10.1007/s13760-020-01328-z (2020).

23 Stenson, P. D. *et al.* The Human Gene Mutation Database (HGMD) and its exploitation in the fields of personalized genomics and molecular evolution. *Curr Protoc Bioinformatics* **Chapter 1**, Unit1 13, doi:10.1002/0471250953.bi0113s39 (2012).

24 Yin, X. *et al.* Identification of a de novo fetal variant in osteogenesis imperfecta by targeted sequencing-based noninvasive prenatal testing. *J Hum Genet* **63**, 1129-1137, doi:10.1038/s10038-018-0489-9 (2018).

25 Zhang, J. *et al.* Non-invasive prenatal sequencing for multiple Mendelian monogenic disorders using circulating cell-free fetal DNA. *Nat Med* **25**, 439-447, doi:10.1038/s41591-018-0334-x (2019).

26 Lv, W. *et al.* Noninvasive Prenatal Testing for Wilson Disease by Use of Circulating Single-Molecule Amplification and Resequencing Technology

(cSMART). *Clinical Chemistry* **61**, 172-181, doi:10.1373/clinchem.2014.229328 (2015).

27 Cutts, A. *et al.* A method for noninvasive prenatal diagnosis of monogenic autosomal recessive disorders. *Blood* **134**, 1190-1193, doi:10.1182/blood.2019002099 (2019).

28 Lun, F. M. *et al.* Noninvasive prenatal diagnosis of monogenic diseases by digital size selection and relative mutation dosage on DNA in maternal plasma. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 19920-19925, doi:10.1073/pnas.0810373105 (2008).

29 Lo, Y. M. *et al.* Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci Transl Med* **2**, 61ra91, doi:10.1126/scitranslmed.3001720 (2010).

30 Vermeulen, C. *et al.* Sensitive Monogenic Noninvasive Prenatal Diagnosis by Targeted Haplotyping. *Am J Hum Genet* **101**, 326-339, doi:10.1016/j.ajhg.2017.07.012 (2017).

31 Suciu, I. D., Toader, O. D., Galeva, S. & Pop, L. Non-Invasive Prenatal Testing beyond Trisomies. *Journal of medicine and life* **12**, 221-224, doi:10.25122/jml-2019-0053 (2019).

32 Brady, P. *et al.* Clinical implementation of NIPT - technical and biological challenges. *Clin Genet* **89**, 523-530, doi:10.1111/cge.12598 (2016).

33 Renga, B. Non invasive prenatal diagnosis of fetal aneuploidy using cell free fetal DNA. *Eur J Obstet Gynecol Reprod Biol* **225**, 5-8, doi:10.1016/j.ejogrb.2018.03.033 (2018).

34 Huang, W., Li, L., Myers, J. R. & Marth, G. T. ART: a next-generation sequencing read simulator. *Bioinformatics* **28**, 593-594, doi:10.1093/bioinformatics/btr708 (2012).

35 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

36 van Beek, D. M. *et al.* Comparing methods for fetal fraction determination and quality control of NIPT samples. *Prenat Diagn* **37**, 769-773, doi:10.1002/pd.5079 (2017).

37 Xu, H. *et al.* Informative priors on fetal fraction increase power of the noninvasive prenatal screen. *Genet Med* **20**, 817-824, doi:10.1038/gim.2017.186 (2018).

38 R Core Team. R: A Language and Environment for Statistical Computing. (2020).

Figures

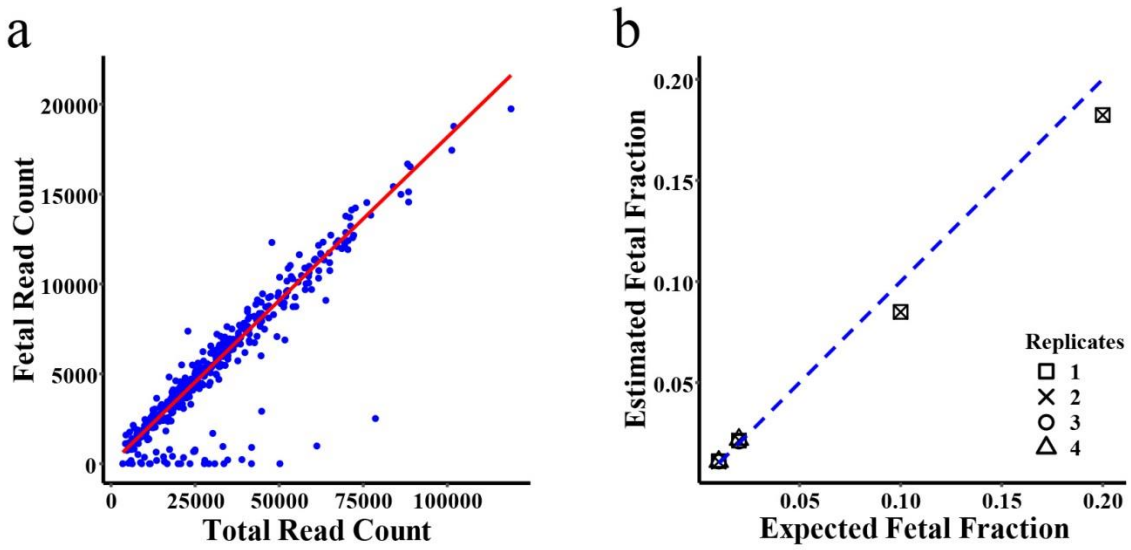


Fig. 1: fetal fraction estimation. a. fetal read count and total read count were estimated for each polymorphic site, and a robust linear regression line was fitted (red line, model $y=\beta x+0$) for each sample of the replication dataset, followed by the estimation of fetal fraction as the model coefficient (β). A representative sample was plotted. b. genomic DNAs from two individuals were mixed at different ratios, and then library or sequencing level replicates were prepared and sequenced for each sample. The expected and estimated fetal fractions were plotted (blue line: $y=x$).

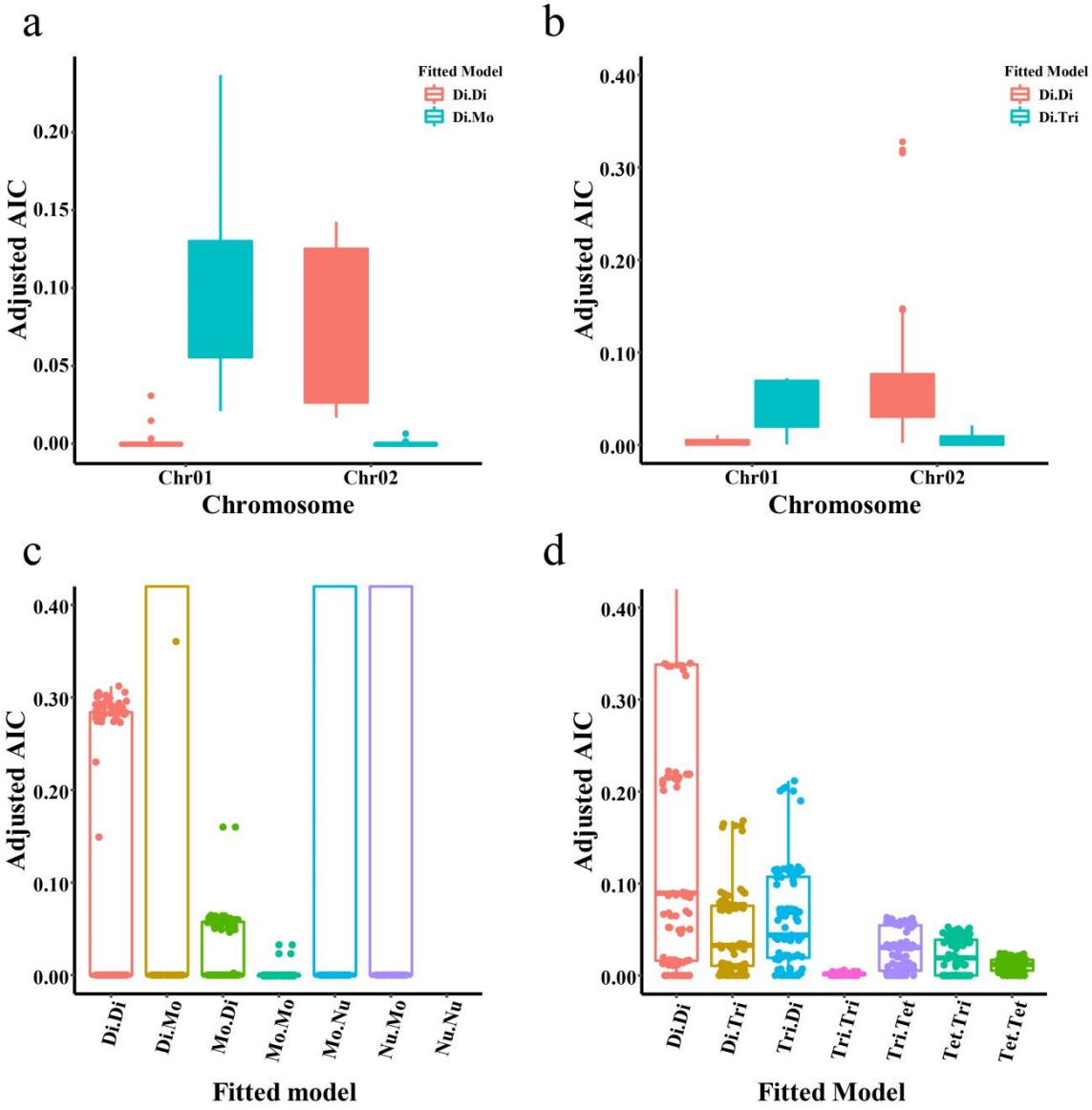


Fig. 2: detection of chromosomal or subchromosomal abnormality. Samples with chromosomal or subchromosomal abnormalities were simulated. Each polymorphic site was tested against all possible chromosomal/subchromosomal models, and the overall fitness for each model was plotted. a. overall fitness of all polymorphic sites to both disomy-disomy (Di.Di) and disomy-monosomy (Di.Mo) models for each chromosome (Chr01: disomy-disomy; Chr02: disomy-monosomy). b. overall fitness of all polymorphic sites to both disomy-disomy (Di.Di) and disomy-trisomy (Di.Tri) models for each chromosome (Chr01: disomy-disomy; Chr02: disomy-trisomy). c. partial enlarged drawings of overall fitted results for a simulated monosomy-monosomy chromosome. d. partial enlarged drawings of overall fitted results for a simulated trisomy-trisomy chromosome. Mo:monosomy. Di: disomy. Tri: trisomy. Nu: nullisomy.

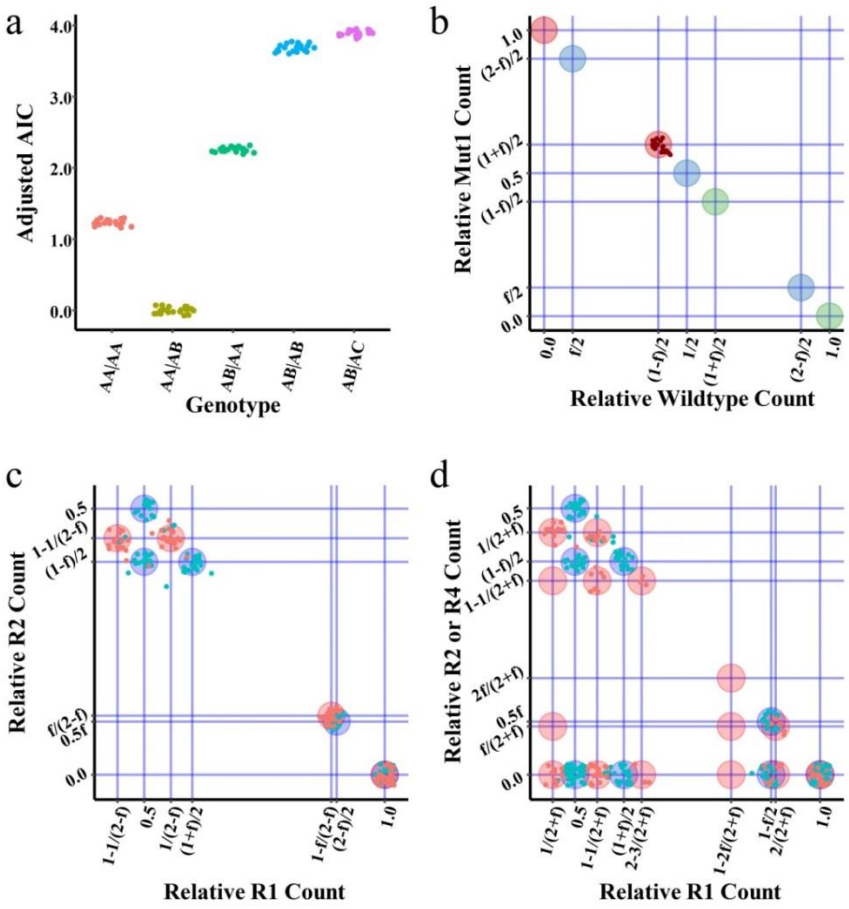


Fig. 3: detection of genetic abnormalities by graphical analysis. a. detection of short genetic variation by allelic goodness-of-fit analysis. The target site with library-level replicates was tested against all possible genotype models and the results plotted. According to the plot, AA|AB genotype was the best fit. Further analysis showed that allele A was mutant and allele B was wildtype, then the target was estimated to be a homozygous mutant-mutant for the mother and a heterozygous wildtype-mutant for the fetus. b. detection of short genetic variation by allelic distribution plot. For a representative two-allele target site with library-level replicates, the most abundant mutant allele's relative count was plotted against the wild type one. According to the cluster position, the target was estimated to be a heterozygous wildtype-mutant for the mother and a homozygous mutant-mutant for the fetus. c. detection of fetal monosomy. Relative allelic counts for polymorphic sites on the reference chromosome (blue) and the target chromosome (red) were plotted for a representative sample. From the characteristic cluster positions, the target chromosome was estimated to be normal for the mother but monosomy for the fetus. d. detection of fetal trisomy. Relative allelic counts for polymorphic sites on the reference chromosome (blue) and the target chromosome (red) were plotted for a representative sample. From the characteristic cluster positions, the target chromosome was estimated to be normal for the mother but trisomy for the fetus.

Supplementary data and figures

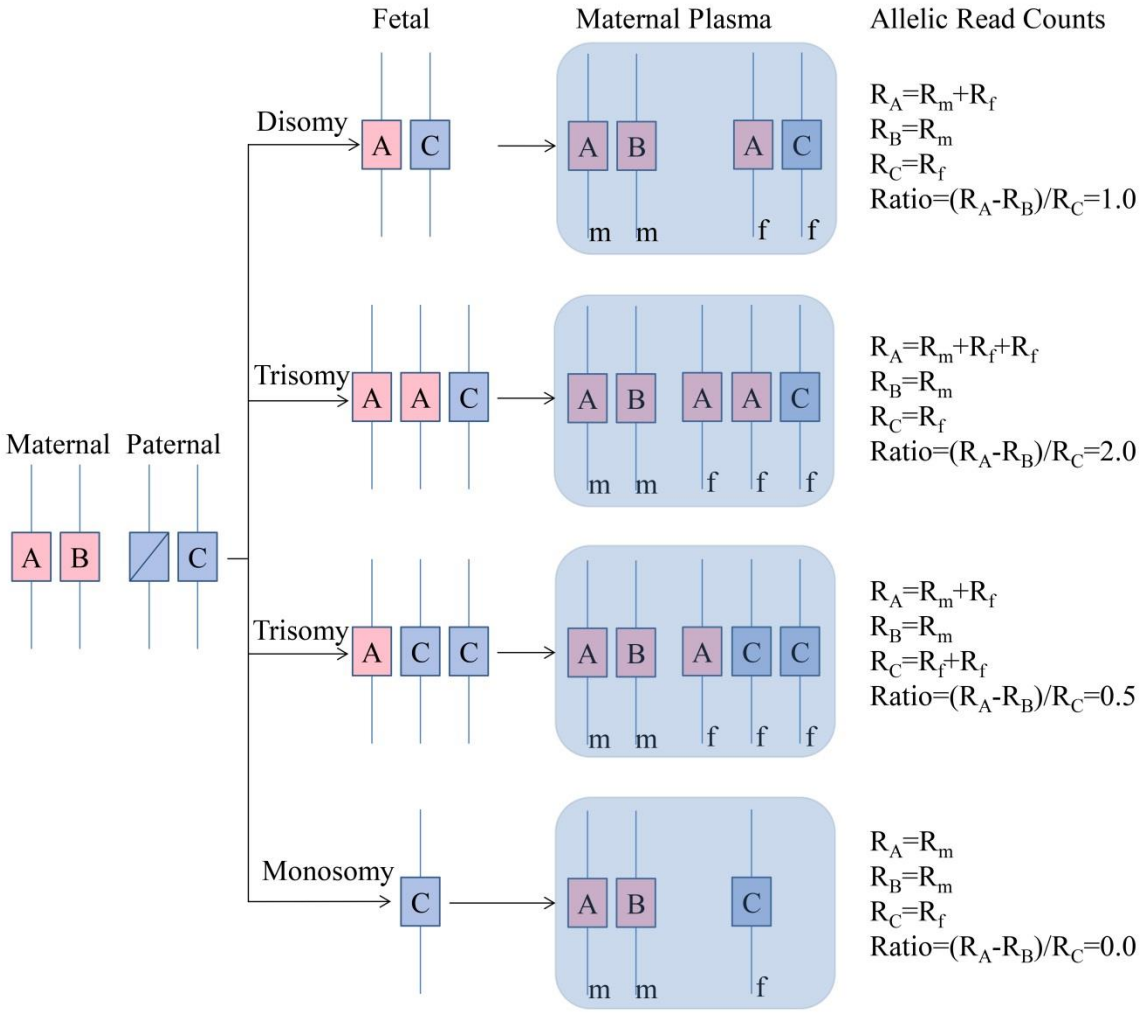


Fig. S1: imbalance of allelic read counts for a polymorphic site from a diploid mother with an aneuploidy fetus inheriting a distinct allele from the father. A, B and C: distinct alleles for a polymorphic site; m and f: maternal and fetal genomic material; R_A, R_B and R_C: allelic read counts for alleles A, B and C, respectively.

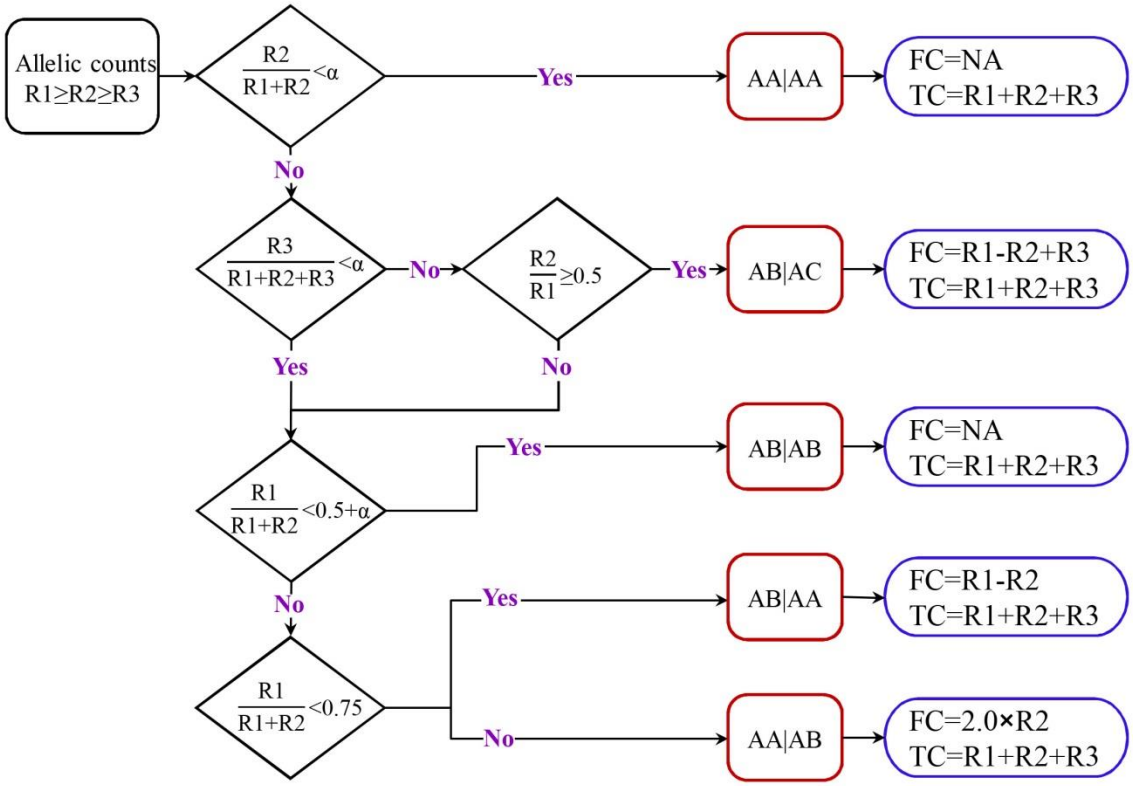


Fig. S2: estimating the maternal-fetal genotype of a polymorphic site using its allelic read counts. R1, R2 and R3: allelic read counts in descending order; α: background threshold. A, B and C are distinct alleles for each polymorphic site, and the portion before the vertical bar denotes the maternal genotype and the part after the vertical bar denotes the fetal genotype. FC: estimated reads count amplified from fetal genetic materials (Fetal Reads); TC: total reads count amplified from both maternal and fetal genetic materials (Total Reads).

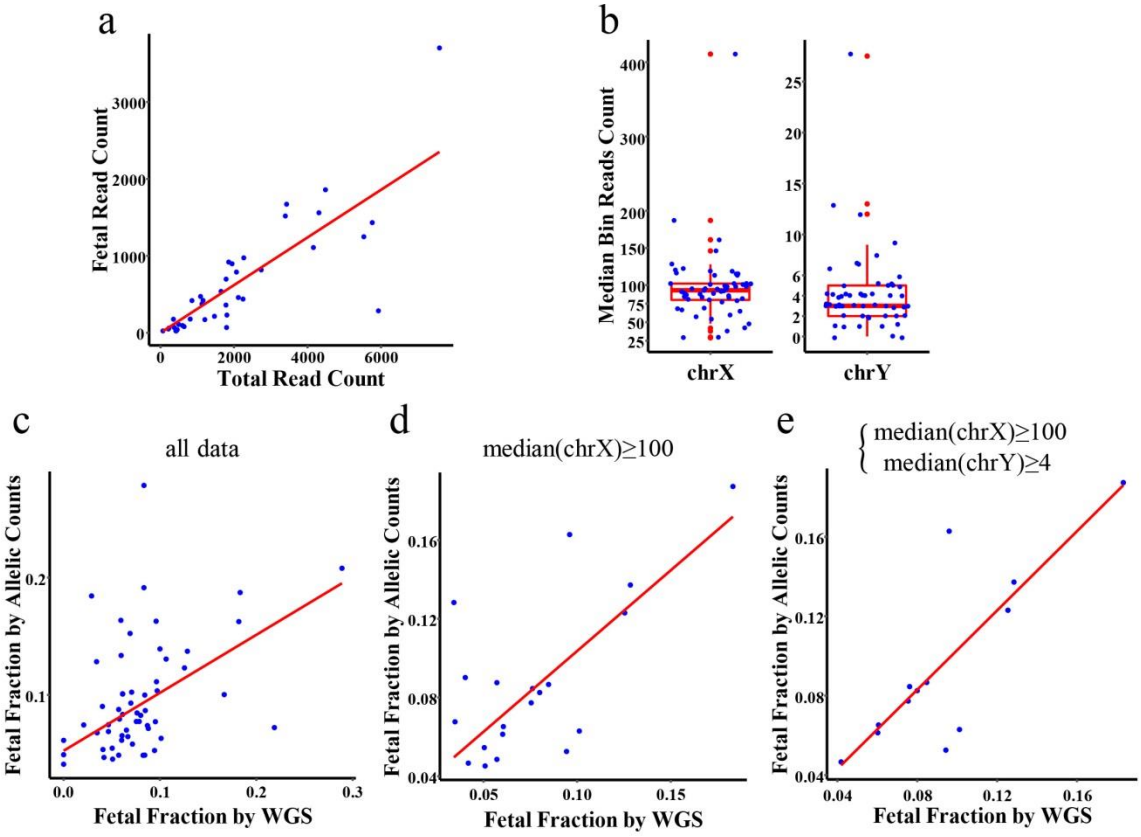


Fig. S3: fetal fraction estimation using allelic read counts or whole genome sequencing. a. for each sample in the insertion/deletion polymorphism dataset, fetal and total read counts were estimated for each polymorphic site, and a robust linear regression line (red line, model $y=\beta x+0$) was fitted followed with the estimation of fetal fraction as the model coefficient (β). A representative plot was shown. b. median bin read counts for WGS dataset. c-e: fetal fractions were estimated for each sample by both allelic read counts and WGS methods, and their relationship was plotted (red line is the fitted regression line $y\sim x$). WGS: whole genome sequencing.

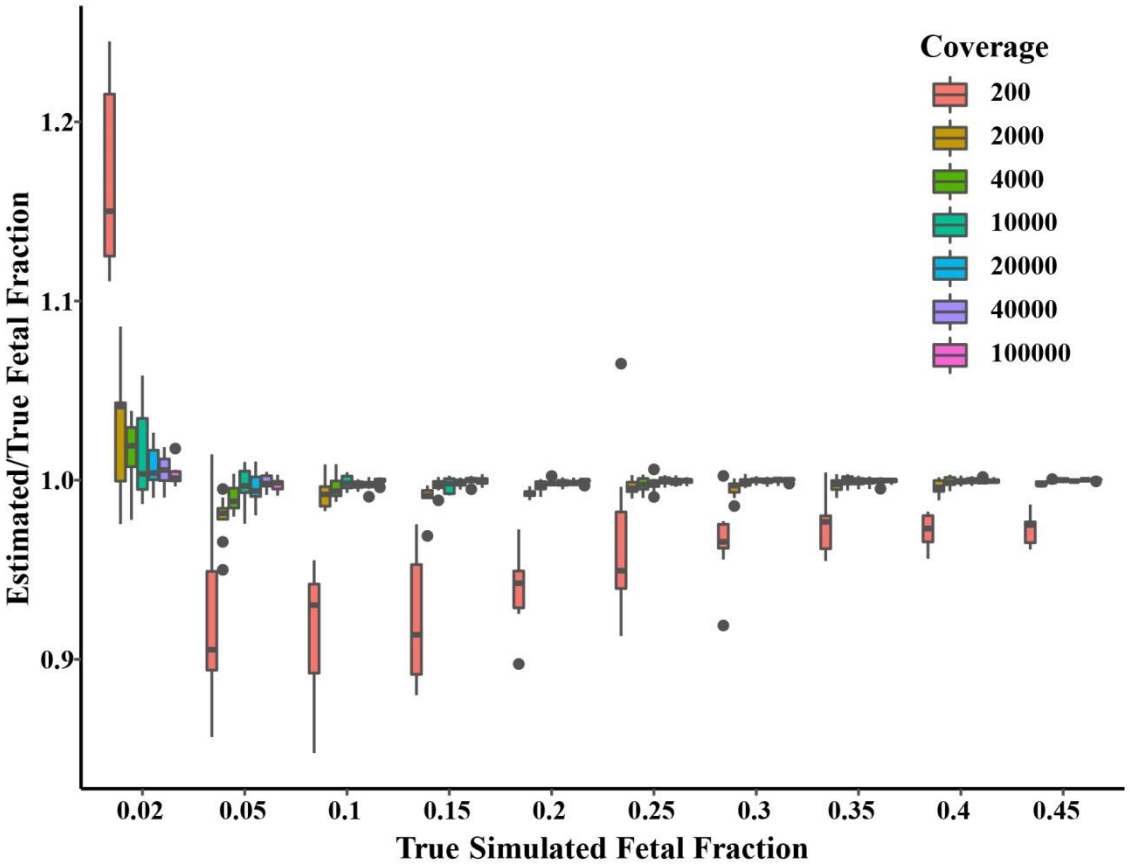


Fig. S4: estimation accuracy for fetal fractions of simulated samples. One hundred samples were simulated for each sequencing coverage, and 100 polymorphic sites were simulated for each sample. At each polymorphic site, allelic sequences for one of the five disomic-disomic genotypes were randomly generated with different fetal fractions. Fetal fraction for each sample was estimated using allelic reads counts. The ratio of the estimated fetal fraction to the true fetal fraction (the simulated value) was plotted and grouped by sequencing coverage.

a

1. Sort all allelic read counts in descending order and label as O_i
2. Calculate total counts $T = \sum O_i$
3. List all possible genotypes and calculate AIC statistics for each genotype as follows:
   (a) Calculate expected allelic counts using fetal fraction (f) for the specific genotype model and label as E_i in descending order;
   (b) Perform G test of goodness-of-fit and calculate AIC as follows:
   $G = 2 \times \sum \left[O_i \times \ln\left(\frac{O_i}{E_i}\right) \right]$
   $AIC = G - 2 \times df$
4. Sort all AICs in ascending order
5. Estimate genotype of the polymorphic site as the one with the minimal AIC and calculate its AIC statistics as follows:
   $\Delta AIC = AIC_{2nd\ minimal} - AIC_{minimal}$
   $adjusted\ AIC = AIC/f/T$
   $adjusted\ \Delta AIC = \Delta AIC/f/T$

b

1. List all possible maternal-fetal chromosomal models for the target chromosome
2. For each polymorphic site, using G test of goodness-of-fit to test against all possible genotypes for each model and select the best fitted genotype for each model, and then calculate its corresponding AIC
3. Analyze AIC statistics for all polymorphic sites on the target chromosome for each model and select the model that shows overall best fit for all polymorphic sites as the estimated maternal-fetal chromosomal model.

c

Expected counts for disomy-disomy model

Genotype	Expected Relative Allelic Counts		
	ER1	ER2	ER3
AA|AA	1-2α	α	α
AA|AB	(1-α)(2-f)/2	(1-α)f/2	α
AB|AA	(1-α)(1+f)/2	(1-α)(1-f)/2	α
AB|AB	(1-α)/2	(1-α)/2	α
AB|AC	1/2	(1-f)/2	f/2

d

Expected relative allelic counts for disomy-monosomy model

Genotype	Expected Relative Allelic Counts		
	ER1	ER2	ER3
AA|AØ	1-2α	α	α
AA|BØ	2(1-α)(1-f)/(2-f)	(1-α)f/(2-f)	α
AB|AØ	(1-α)/(2-f)	(1-α)(1-f)/(2-f)	α
AB|CØ	(1-f)/(2-f)	(1-f)/(2-f)	f/(2-f)

e

Expected relative allelic counts for disomy-trisomy model

Genotype	Expected Relative Allelic Counts			
	ER1	ER2	ER3	ER4
AA|AAA	1-3α	α	α	α
AA|AAB	2(1-2α)/(2+f)	(1-2α)f/(2+f)	α	α
AA|ABB	(1-2α)(2-f)/(2+f)	2(1-2α)f/(2+f)	α	α
AB|AAA	(1-2α)(1+2f)/(2+f)	(1-2α)(1-f)/(2+f)	α	α
AB|AAB	(1-2α)(1+f)/(2+f)	(1-2α)/(2+f)	α	α
AA|ABC	(1-α)(2-f)/(2+f)	(1-α)f/(2+f)	(1-α)f/(2+f)	α
AB|AAC	(1-α)(1+f)/(2+f)	(1-α)(1-f)/(2+f)	(1-α)f/(2+f)	α
AB|ABC	(1-α)/(2+f)	(1-α)/(2+f)	(1-α)f/(2+f)	α
AB|ACC	(1-α)/(2+f)	(1-α)/(2+f)	2(1-α)f/(2+f)	α
AB|ACD	1/(2+f)	(1-f)/(2+f)	f/(2+f)	f/(2+f)

Fig. S5: maternal-fetal genotype estimation. a. steps to estimate the maternal-fetal genotype for each polymorphic site. b. steps to estimate the maternal-fetal chromosomal model for each target chromosome or subchromosomal fragment. c: expected relative allelic counts for each polymorphic site on a maternal-fetal disomy-disomy chromosome. d: expected relative allelic counts for each polymorphic site on a maternal-fetal disomy-monosomy chromosome. e: expected relative allelic counts for each polymorphic site on a maternal-fetal disomy-trisomy chromosome. α: noise background. Expected allelic count was calculated as the product of total counts (T) and the corresponding expected relative allelic count.

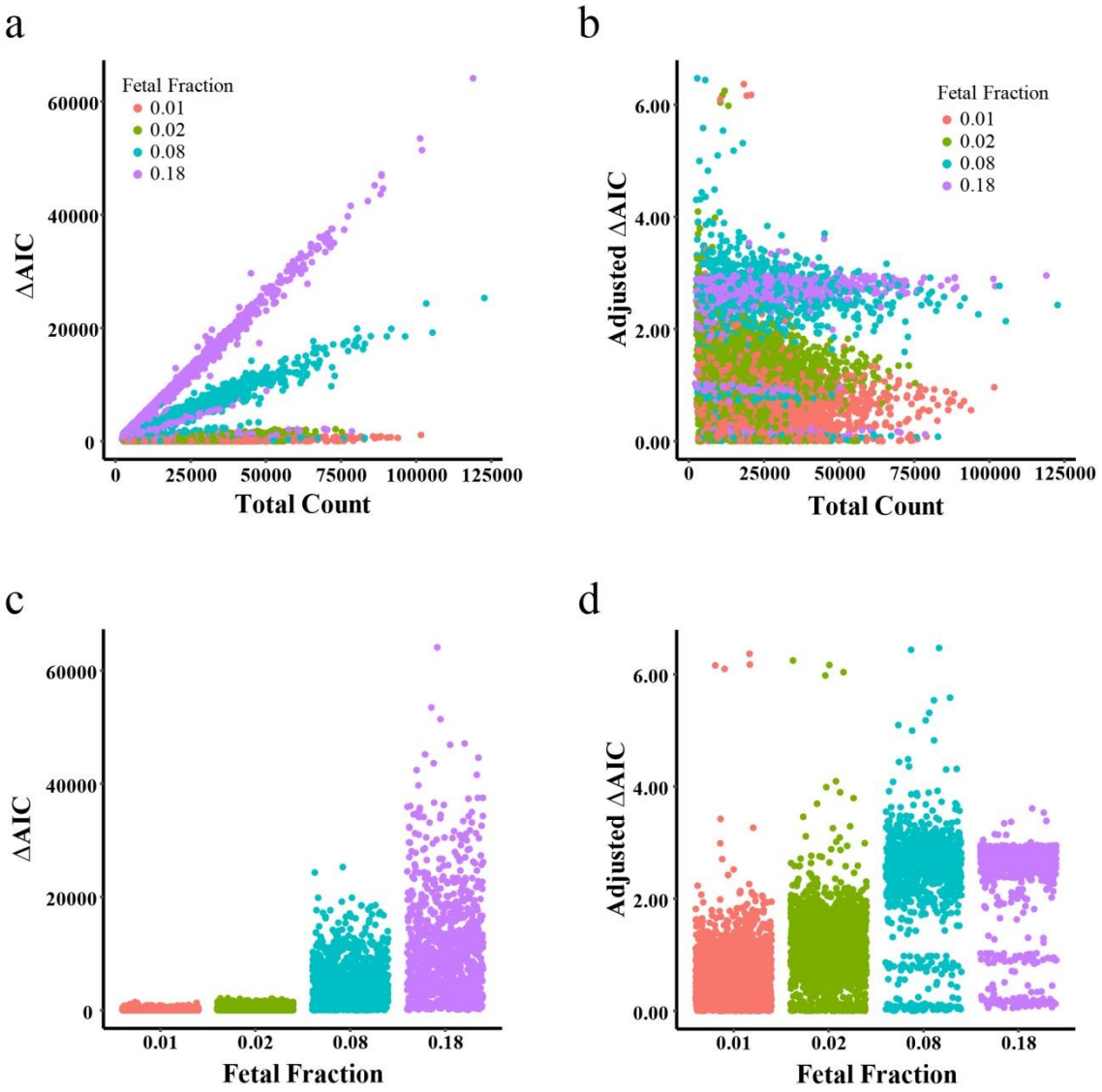


Fig. S6: influence of fetal fraction and total allelic read count on ΔAIC. Fetal fraction was estimated for each sample in the replicates dataset and rounded to the second decimal place. a. absolute ΔAIC was calculated for each polymorphic site of the replicate samples using the disomic-disomic model, plotted against total allelic read count and grouped by the estimated fetal fraction. b: absolute adjusted ΔAICs were calculated for the replicate samples and plotted against total allelic read counts. c: absolute ΔAICs were calculated for the replicate samples and plotted against fetal fractions. d: absolute adjusted ΔAICs were calculated for the replicate samples and plotted against fetal fractions.

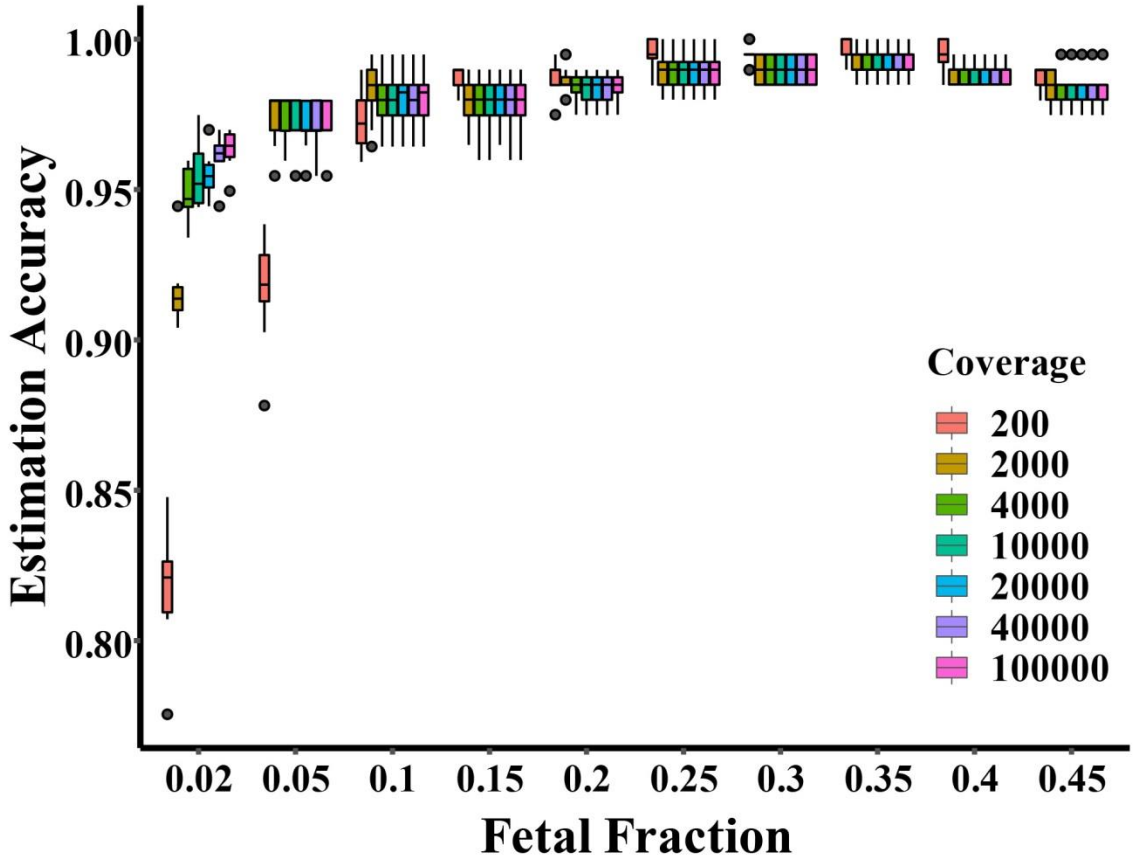


Fig. S7: influence of fetal fraction on maternal-fetal genotype estimation. Sequencing reads were simulated for samples with different fetal fractions and different sequencing coverage, and fetal fraction was estimated for each sample followed with genotype estimation for each polymorphic site. Estimation accuracy was calculated as the ratio of the number of correctly estimated genotypes to the total number of all polymorphic sites, and then plotted against fetal fraction grouped by sequencing coverage.

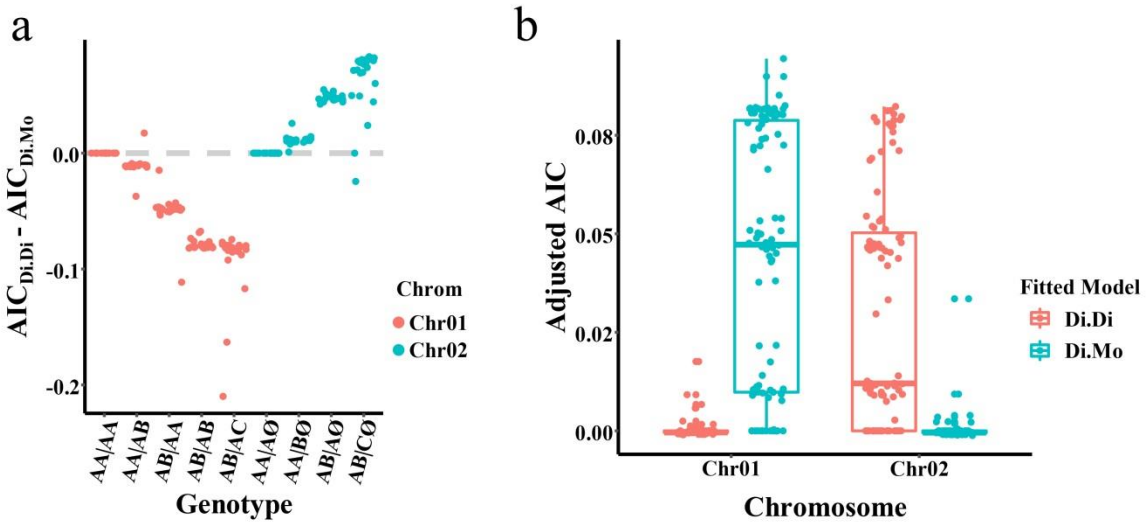


Fig. S8: detection of fetal monosomy. Two pairs of chromosomes were simulated and one (Chr01) was disomy-disomy and the other one (Chr02) was disomy-monosomy. For each polymorphic site, two minimal adjusted AIC values were calculated, one ($AIC_{Di.Di}$) was for fitting genotypes assuming a disomy-disomy model (Di.Di) and the other one ($AIC_{Di.Mo}$) assuming a disomy-monosomy model (Di.Mo). a: ΔAIC for each polymorphic site was calculated as $AIC_{Di.Di}$ - $AIC_{Di.Mo}$ and plotted against its true genotype (simulated genotype). b: each polymorphic site was tested against both the disomy-disomy model (Di.Di) and the disomy-monosomy model (Di.Mo), and the minimal adjusted AICs were plotted for each chromosome.

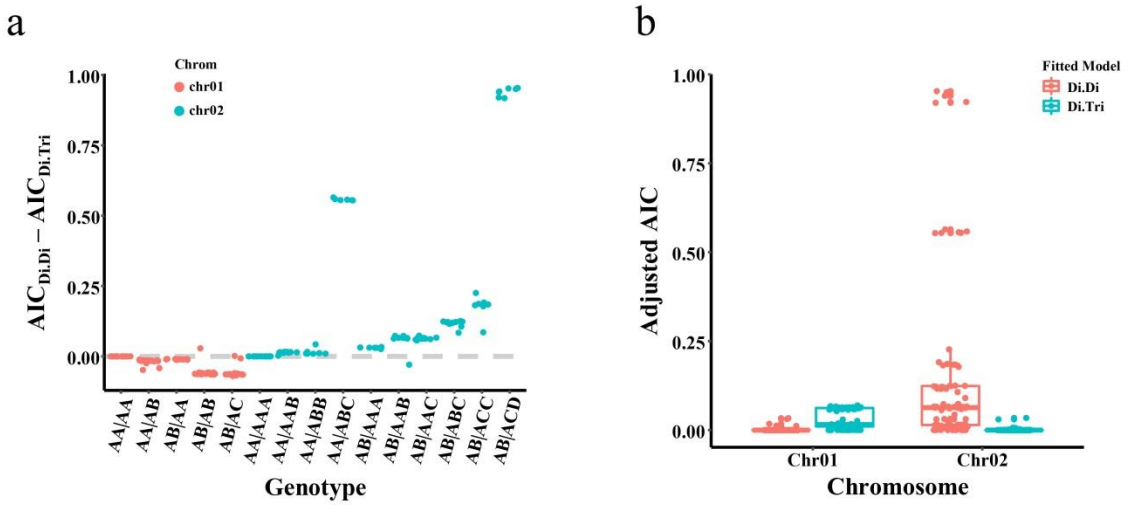


Fig. S9: detection of fetal trisomy. Two pairs of chromosomes were simulated and one (chr01) was disomy-disomy and the other one (chr02) was disomy-trisomy. For each polymorphic site, two minimal adjusted AIC values were calculated, one ($AIC_{Di.Di}$) was for fitting genotypes assuming a disomy-disomy model (Di.Di) and the other one ($AIC_{Di.Tri}$) assuming a disomy-trisomy model (Di.Tri). a: ΔAIC for each polymorphic site was calculated as $AIC_{Di.Di}$ - $AIC_{Di.Tri}$ and plotted against its true genotype (simulated genotype). b: each polymorphic site was tested against both the disomy-disomy model (Di.Di) and the disomy-trisomy model (Di.Tri), and the minimal adjusted AICs were plotted for each chromosome.

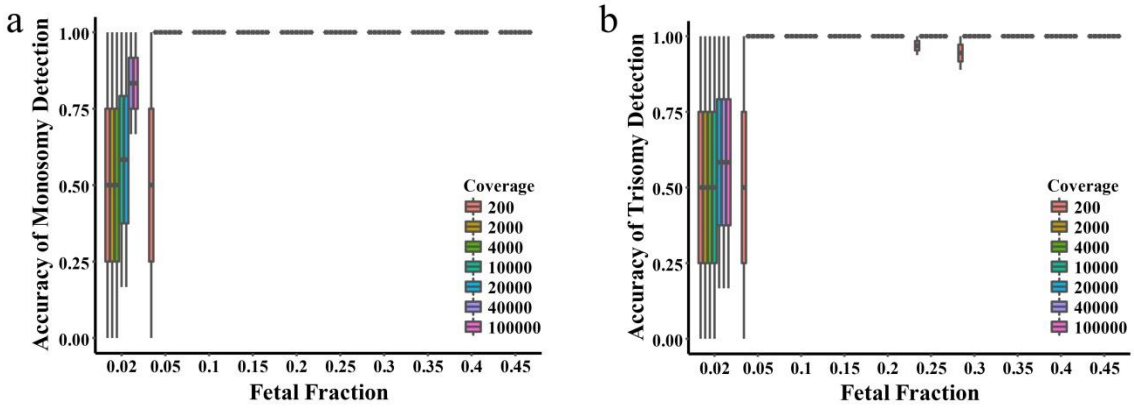


Fig. S10: detecting accuracy for chromosomal aneuploidies. a: in each sample, one disomy-disomy chromosome and one disomy-monosomy chromosome were simulated. Chromosomal aneuploidy for each chromosome in each sample was estimated using overall allelic goodness-of-fit test. One hundred samples with different fetal fractions were simulated for each sequencing coverage. b: in each sample, one disomy-disomy chromosome and one disomy-trisomy chromosome were simulated. Chromosomal aneuploidy for each chromosome in each sample was estimated using overall allelic goodness-of-fit test. One hundred samples with different fetal fractions were simulated for each sequencing coverage.

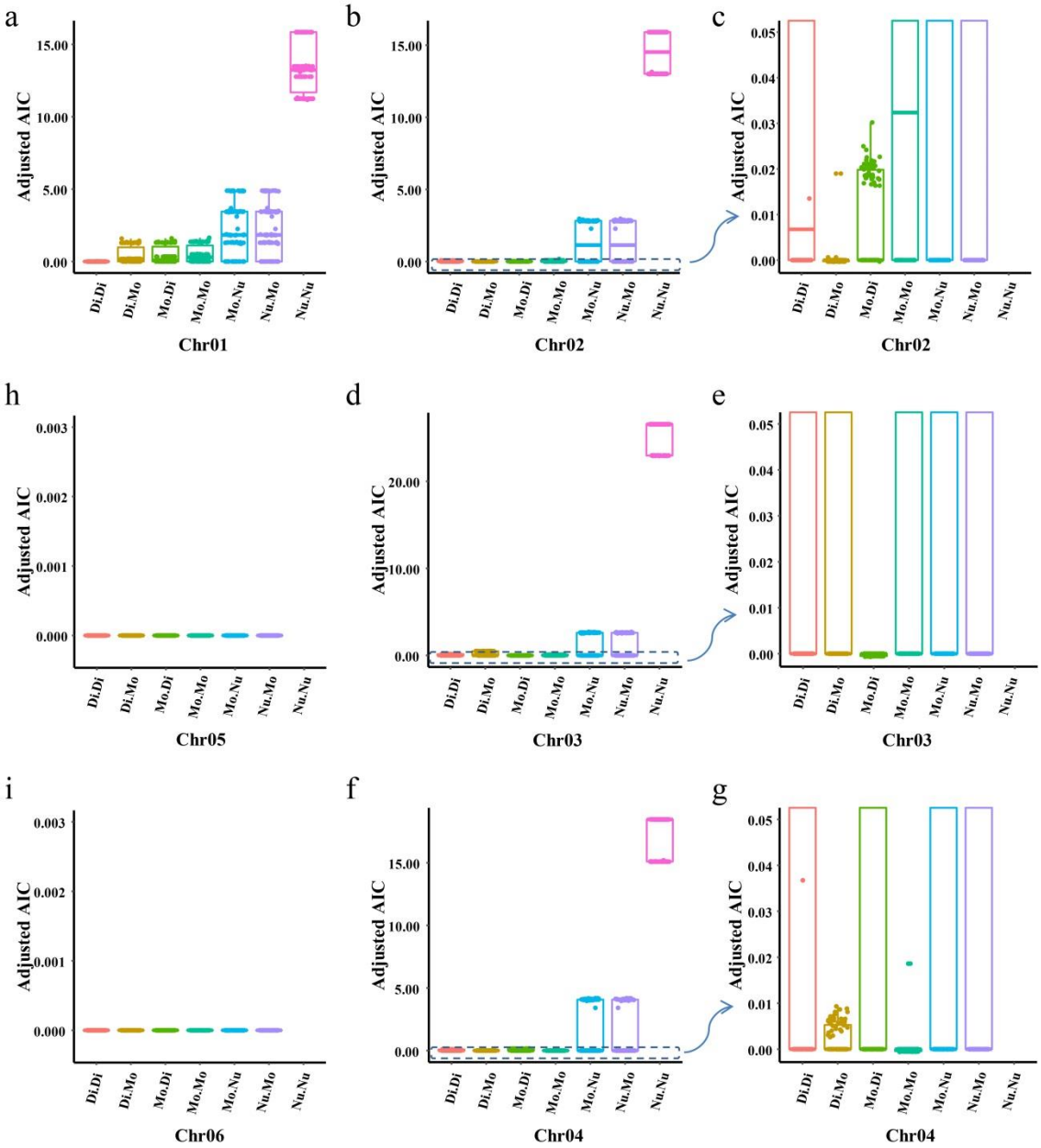


Fig. S11: detection of subchromosomal deletion. Six pairs of chromosomes were simulated with the labels Ch01-Chr06 and their karyotypes were disomy-disomy (Di.Di), disomy-monosomy (Di.Mo), monosomy-disomy (Mo.Di), monosomy-monosomy(Mo.Mo), monosomy-nullisomy (Mo.Nu) and nullisomy-monosomy (Nu.Mo), respectively. Allelic read counts of each polymorphic site was tested against all possible genotypes assuming each one of the seven chromosomal karyotype models (Di.Di, Di.Mo, Mo.Di, Mo.Mo, Mo.Nu, Nu.Mo and Nu.Nu), and the overall best fitted model for all target polymorphic sites was selected for each chromosome. a,b,d,f,h,i: overall fitted results for each chromosome. c,e,g: partial enlarged drawings of b, d and f, respectively.

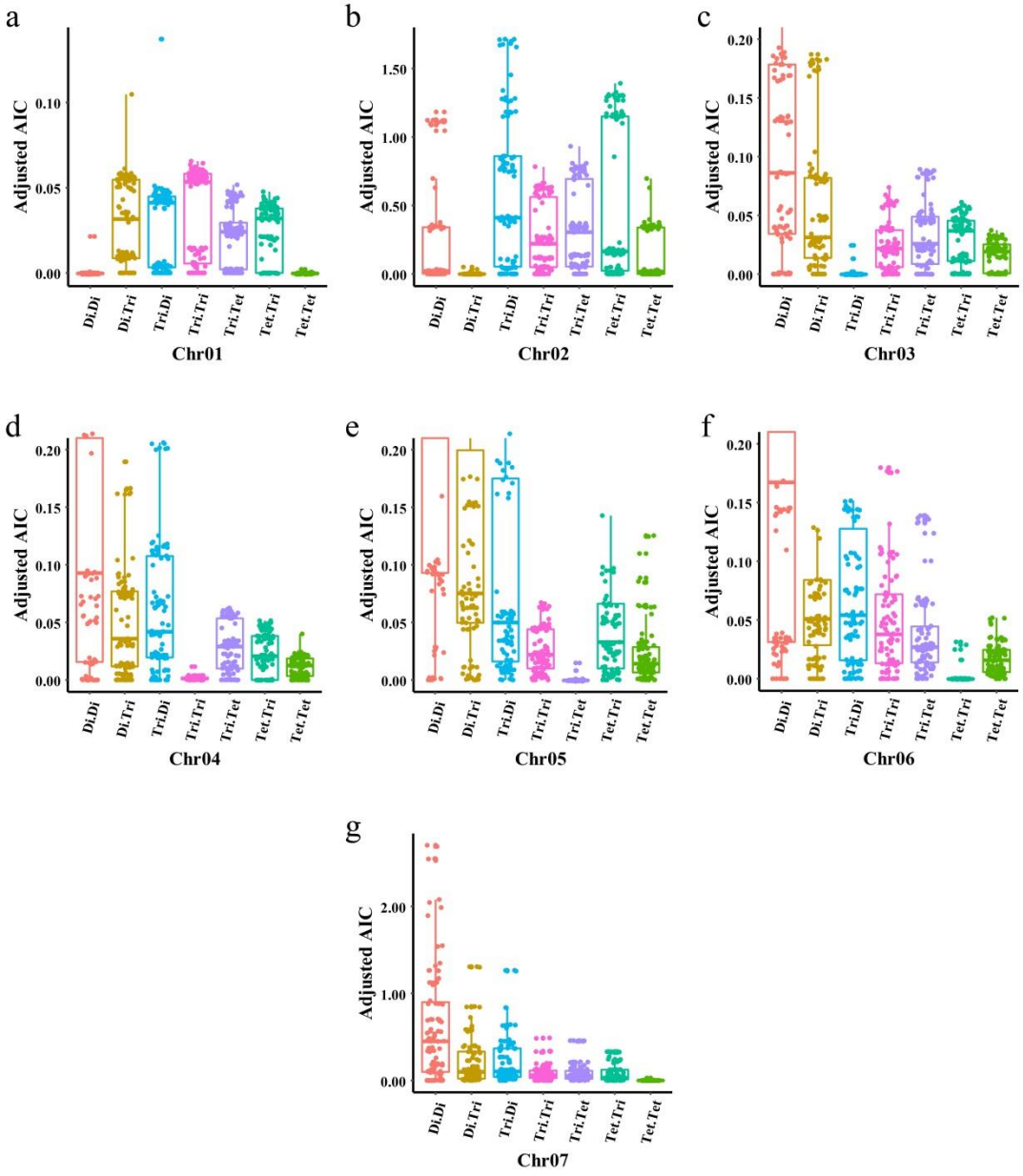


Fig. S12: detection of subchromosomal duplication. Seven pairs of chromosomes were simulated with the labels Ch01-Chr07 and their karyotypes were disomy-disomy (Di.Di), disomy-trisomy (Di.Tri), trisomy-disomy (Tri.Di), trisomy-trisomy (Tri.Tri), trisomy-tetrasomy (Tri.Tet), tetrasomy-trisomy (Tet.Tri) and tetrasomy-tetrasomy(Tet.Tet), respectively. Allelic read counts of each polymorphic site was tested against all possible genotypes assuming each one of the seven chromosomal karyotype models (Di.Di, Di.Tri, Tri.Di, Tri.Tri, Tri.Tet, Tet.Tri and Tet.Tet), and the overall best fitted model for all target polymorphic sites was selected for each chromosome. a,b,g: overall fitted results for chromosomes Chr01, Chr02 and Chr07, respectively. c,d,e,f: partial enlarged drawings of overall fitted results for chromosomes Chr03, Chr04, Chr05 and Chr06, respectively.

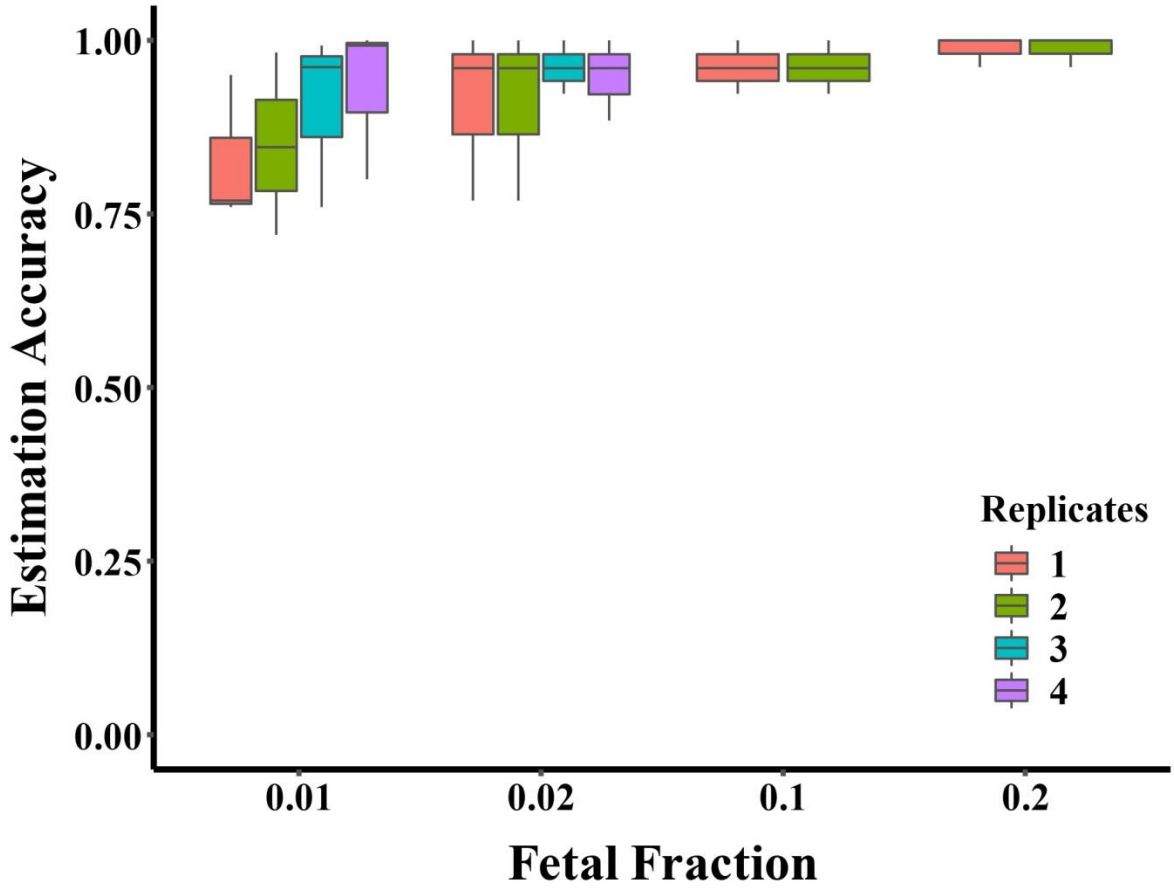


Fig. S13: genotype estimation accuracy for the replication dataset. Genotype was estimated for each polymorphic site using its allelic read counts for each sample in the replication dataset. Estimation accuracy was calculated as the ratios of the number of correctly estimated genotypes to the total number of polymorphic sites grouped by different replicates and different fetal fractions. Replicates were labeled as 1 to 4, and ratios for replicates 1 to 4 means that 1 to 4 samples were used to calculate the estimation accuracy, respectively.

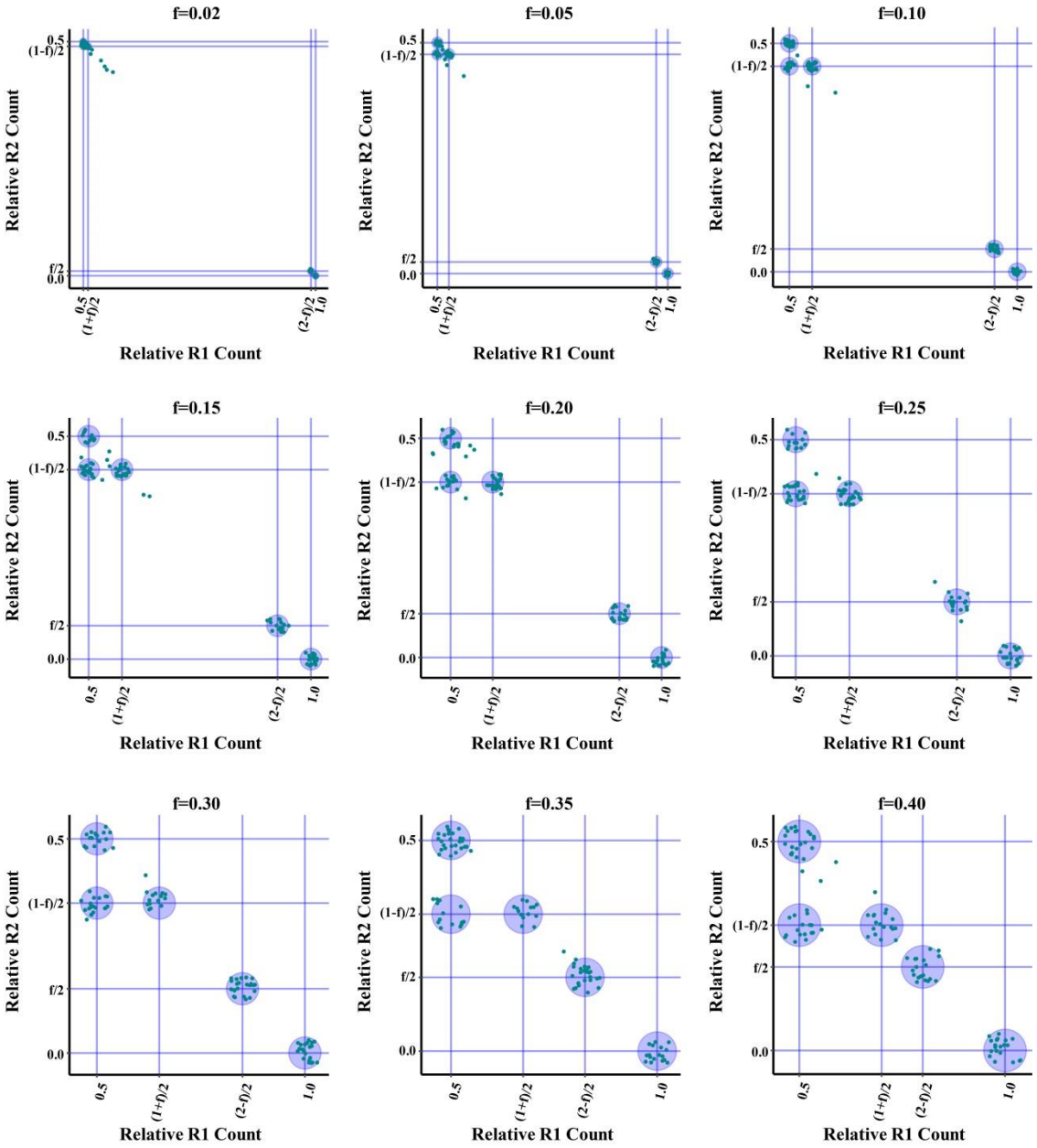
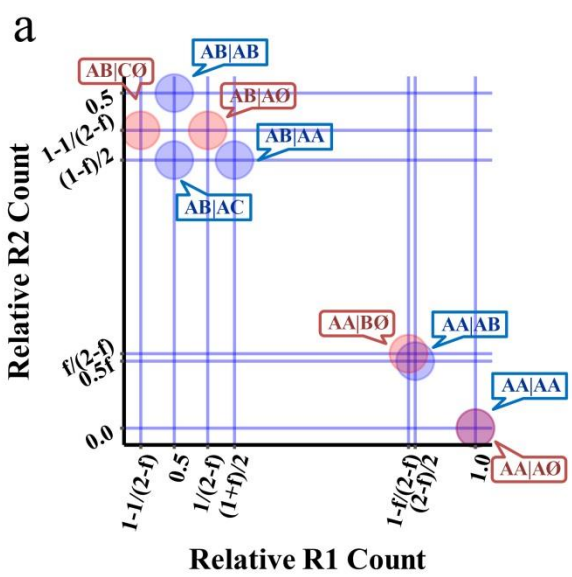


Fig. S14: distribution plots of relative allelic counts. One hundred polymorphic sites on a disomy-disomy chromosome were simulated for each sample. For each polymorphic site in a sample, relative allelic read counts were calculated, and then the relative R2 count was plotted against the relative R1 count. One representative plot was shown for each fetal fraction. f: fetal fraction. Relative R1 Count=R1/(R1+R2+R3) and Relative R2 Count=R2/(R1+R2+R3).

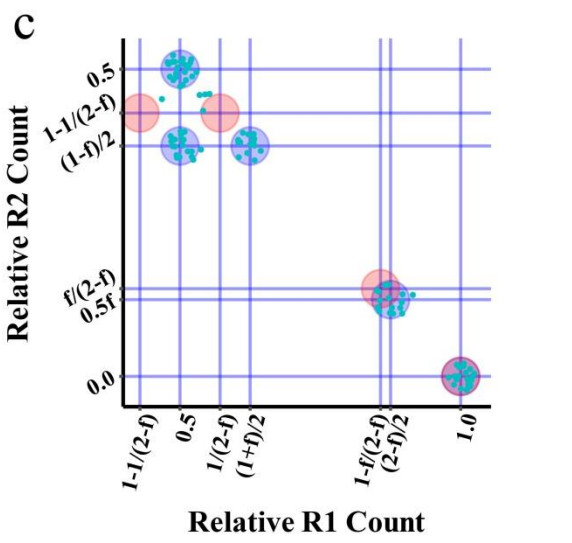


b

Expected Relative Allelic Count

Genotype		RRC1	RRC2
Di.Di	AA|AA	1	0
	AA|AB	(2-f)/2	f/2
	AB|AA	(1+f)/2	(1-f)/2
	AB|AB	1/2	1/2
	AB|AC	1/2	(1-f)/2
Di.Mo	AA|AØ	1	0
	AA|BØ	1-f/(2-f)	f/(2-f)
	AB|AØ	1/(2-f)	1-1/(2-f)
	AB|CØ	1-1/(2-f)	1-1/(2-f)

1. RRC1 (relative R1 count)=R1/(R1+R2+R3)
2. RRC2 (relative R2 count)=R2/(R1+R2+R3)



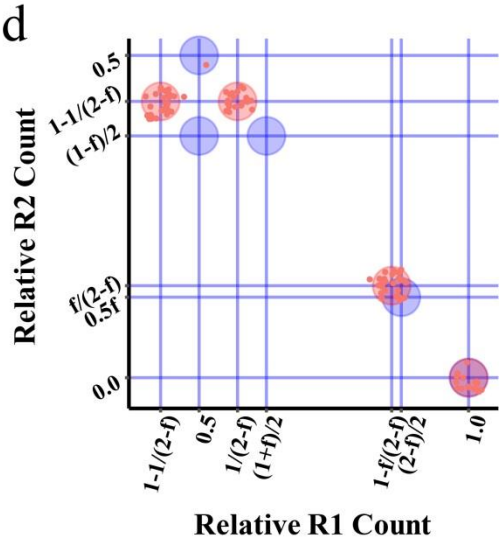
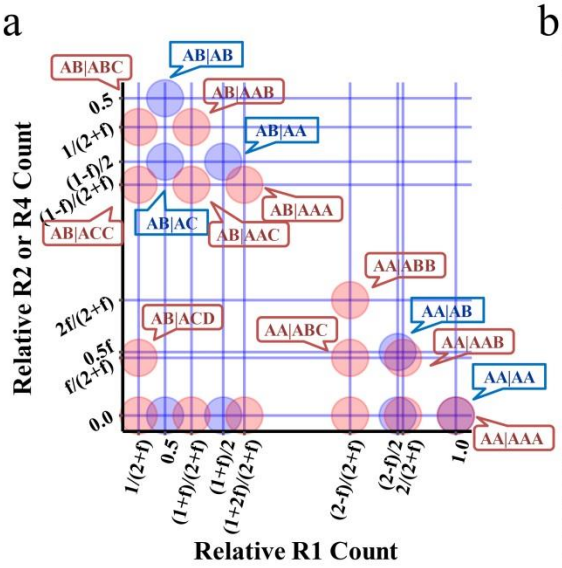


Fig. S15: detection of chromosomal monosomy. a. expected possible positions for polymorphic sites on a normal disomy-disomy chromosome (shaded blue) or on a disomy-monosomy chromosome (shaded red). b. expected relative allelic counts for each polymorphic site. c. relative allelic count plot for polymorphic sites on a representative target chromosome. From the characteristic cluster positions, the target chromosome was estimated to be disomy-disomy. d. relative allelic count plot for polymorphic sites on a representative target chromosome. From the characteristic cluster positions, the target chromosome was estimated to be disomy-monosomy. Di.Di: disomy-disomy. Di.Mo: disomy-monosomy. f: fetal fraction.

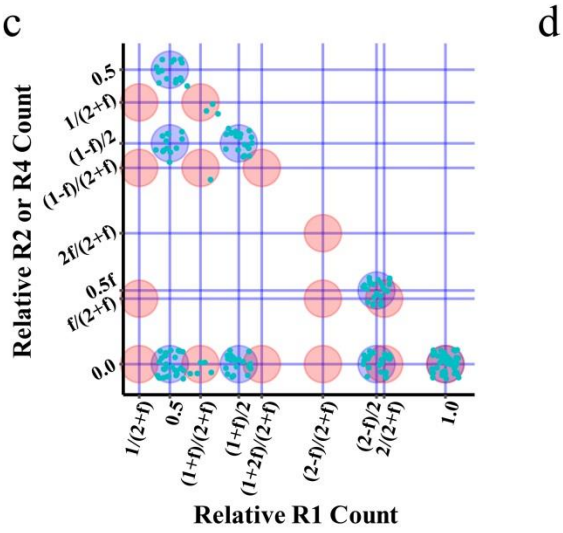


b

Expected Relative Allelic Count

Genotype		RRC1	RRC2	RRC3	RRC4
Di.Di	AA|AA	1	0	0	0
	AA|AB	(2-f)/2	f/2	0	0
	AB|AA	(1+f)/2	(1-f)/2	0	0
	AB|AB	1/2	1/2	0	0
	AB|AC	1/2	(1-f)/2	f/2	0
Di.Tri	AA|AAA	1	0	0	0
	AA|AAB	2/(2+f)	f/(2+f)	0	0
	AA|ABB	(2-f)/(2+f)	2f/(2+f)	0	0
	AB|AAA	(1+2f)/(2+f)	(1-f)/(2+f)	0	0
	AB|AAB	(1+f)/(2+f)	1/(2+f)	0	0
	AA|ABC	(2-f)/(2+f)	f/(2+f)	f/(2+f)	0
	AB|AAC	(1+f)/(2+f)	(1-f)/(2+f)	f/(2+f)	0
	AB|ABC	1/(2+f)	1/(2+f)	f/(2+f)	0
	AB|ACC	1/(2+f)	(1-f)/(2+f)	2f/(2+f)	0
	AB|ACD	1/(2+f)	(1-f)/(2+f)	f/(2+f)	f/(2+f)

1. RRC1 (relative R1 count)=R1/(R1+R2+R3+R4)
2. RRC2 (relative R2 count)=R2/(R1+R2+R3+R4)
3. RRC3 (relative R3 count)=R3/(R1+R2+R3+R4)
4. RRC4 (relative R4 count)=R4/(R1+R2+R3+R4)



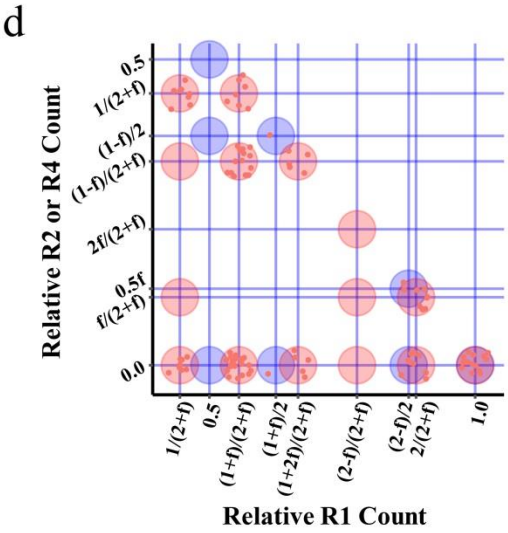
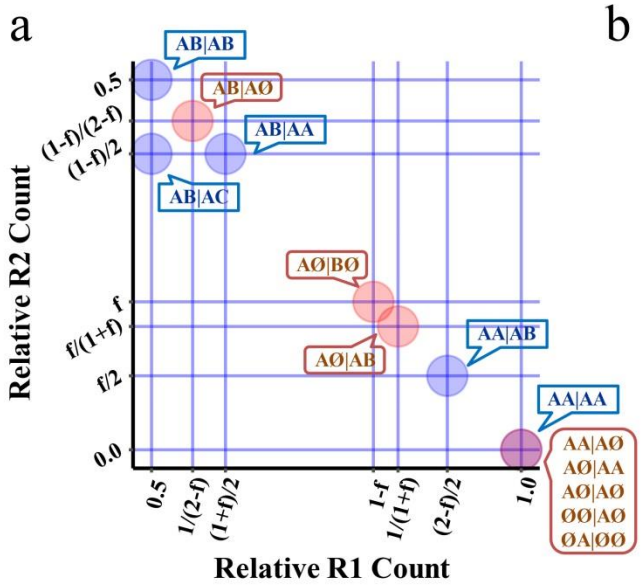


Fig. S16: detection of chromosomal trisomy. a. expected possible positions for polymorphic sites on a normal disomy-disomy chromosome (shaded blue) or on a disomy-trisomy chromosome (shaded red). b. expected relative allelic counts for each polymorphic site. c. relative allelic count plot for polymorphic sites on a representative target chromosome. From the characteristic cluster positions, the target chromosome was estimated to be disomy-disomy. d. relative allelic count plot for polymorphic sites on a representative target chromosome. From the characteristic cluster positions, the target chromosome was estimated to be disomy-trisomy. Di.Di: disomy-disomy. Di.Tri: disomy-trisomy. f: fetal fraction.



b

Expected Relative Allelic Count

Genotype		RRC1	RRC2
Di.Di	AA|AA	1	0
	AA|AB	(2-f)/2	f/2
	AB|AA	(1+f)/2	(1-f)/2
	AB|AB	1/2	1/2
	AB|AC	1/2	(1-f)/2
Di.Mo	AA|AØ	1	0
	AB|AØ	1/(2-f)	(1-f)/(2-f)
Mo.Di	AØ|AA	1	0
	AØ|AB	1/(1+f)	f/(1+f)
Mo.Mo	AØ|AØ	1	0
	AØ|BØ	1-f	f
Nu.Mo	ØØ|AØ	1	0
Mo.Nu	ØA|ØØ	1	0
Nu.Nu	ØØ|ØØ	0	0

1. RRC1 (relative R1 count)=R1/(R1+R2+R3)
2. RRC2 (relative R2 count)=R2/(R1+R2+R3)
3. Only heritable genotypes were listed.

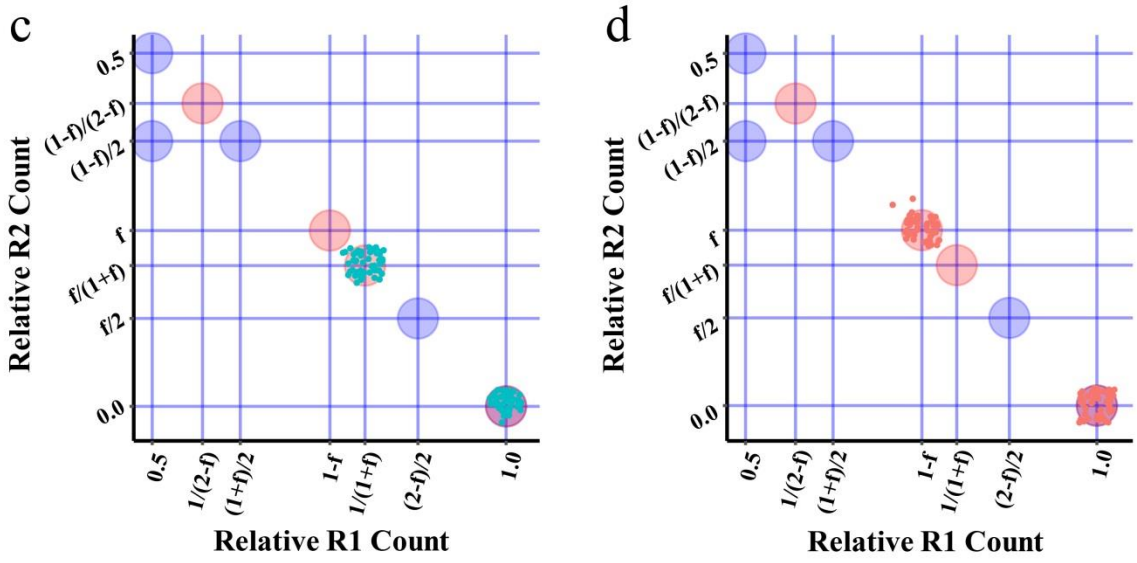


Fig. S17: detection of subchromosomal deletion. a. expected possible positions for polymorphic sites on a normal disomy-disomy chromosome (shaded blue) or on a chromosome with a subchromosomal deletion (shaded red). b. expected relative allelic counts for each polymorphic site. c. relative allelic count plot for polymorphic sites on a representative target chromosome. From the characteristic cluster positions, the target chromosome was estimated to be monosomy-disomy. d. relative allelic count plot for polymorphic sites on a representative target chromosome. From the characteristic cluster positions, the target chromosome was estimated to be monosomy-monosomy. Di.Di: disomy-disomy. Di.Mo: disomy-monosomy. Mo.Di: monosomy-disomy. Mo.Mo: monosomy-monosomy. Nu.Mo: nullisomy-monosomy. Mo.Nu: monosomy-nullisomy. Nu.Nu: nullisomy-nullisomy. f: fetal fraction.

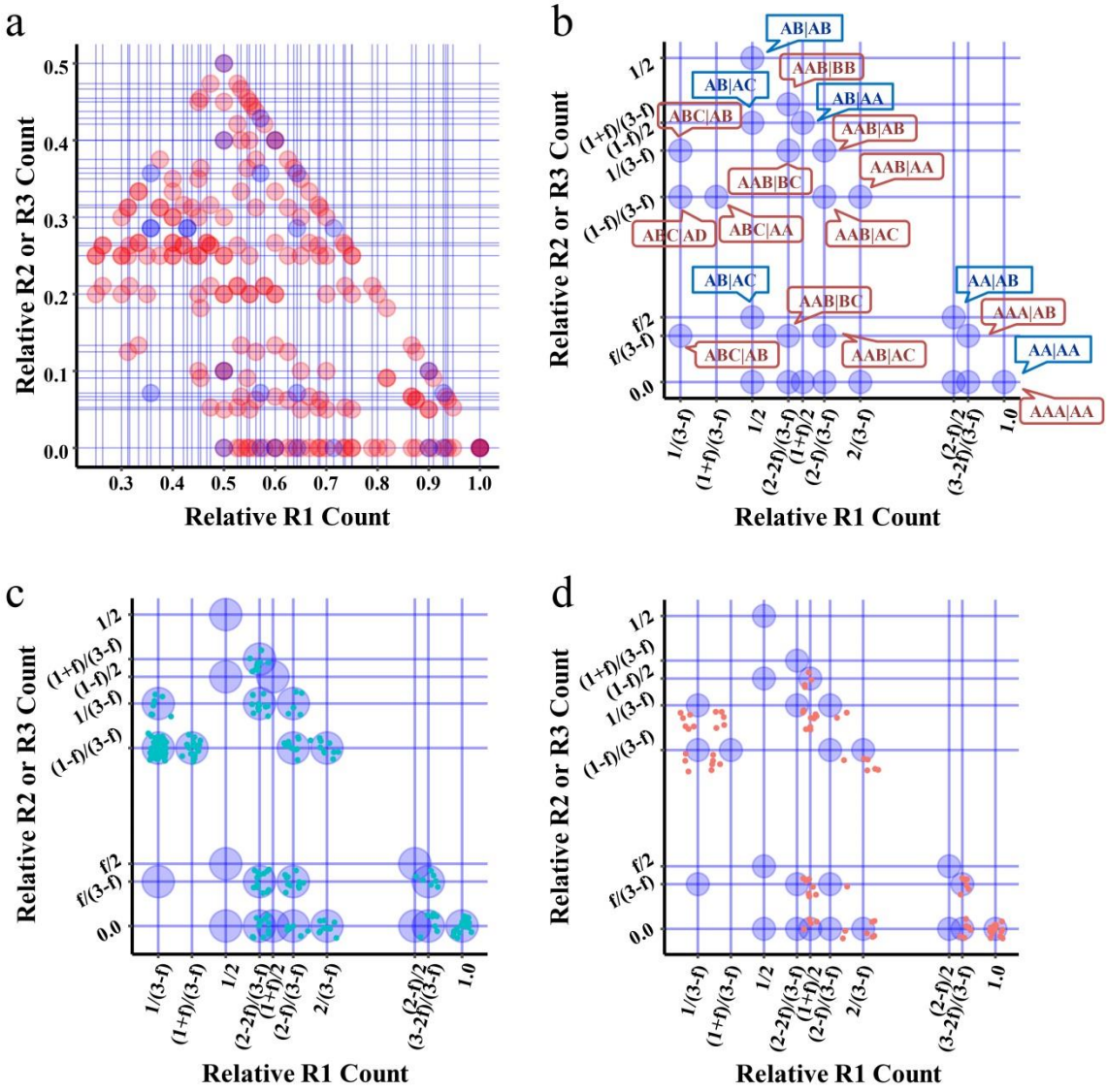


Fig. S18: detection of subchromosomal duplication. a. expected possible positions for polymorphic sites on a normal disomy-disomy chromosome or on a chromosome with subchromosomal duplications (blue: genotype clusters for a chromosome that is normal in fetus. red: genotype clusters for a chromosome that has one or more microduplications in fetus). b. expected possible positions for polymorphic sites on a chromosome that is disomy for the fetus. c. relative allelic count plot for polymorphic sites on a representative target chromosome. From the characteristic cluster positions, the target chromosome was estimated to be normal for the fetus but abnormal for the mother (trisomy-disomy specifically). d. relative allelic count plot for polymorphic sites on a representative target chromosome. As there were allelic clusters not in the expected positions for a normal fetus, either the fetus was abnormal for microduplications or the true and correct model was not included in the analysis.

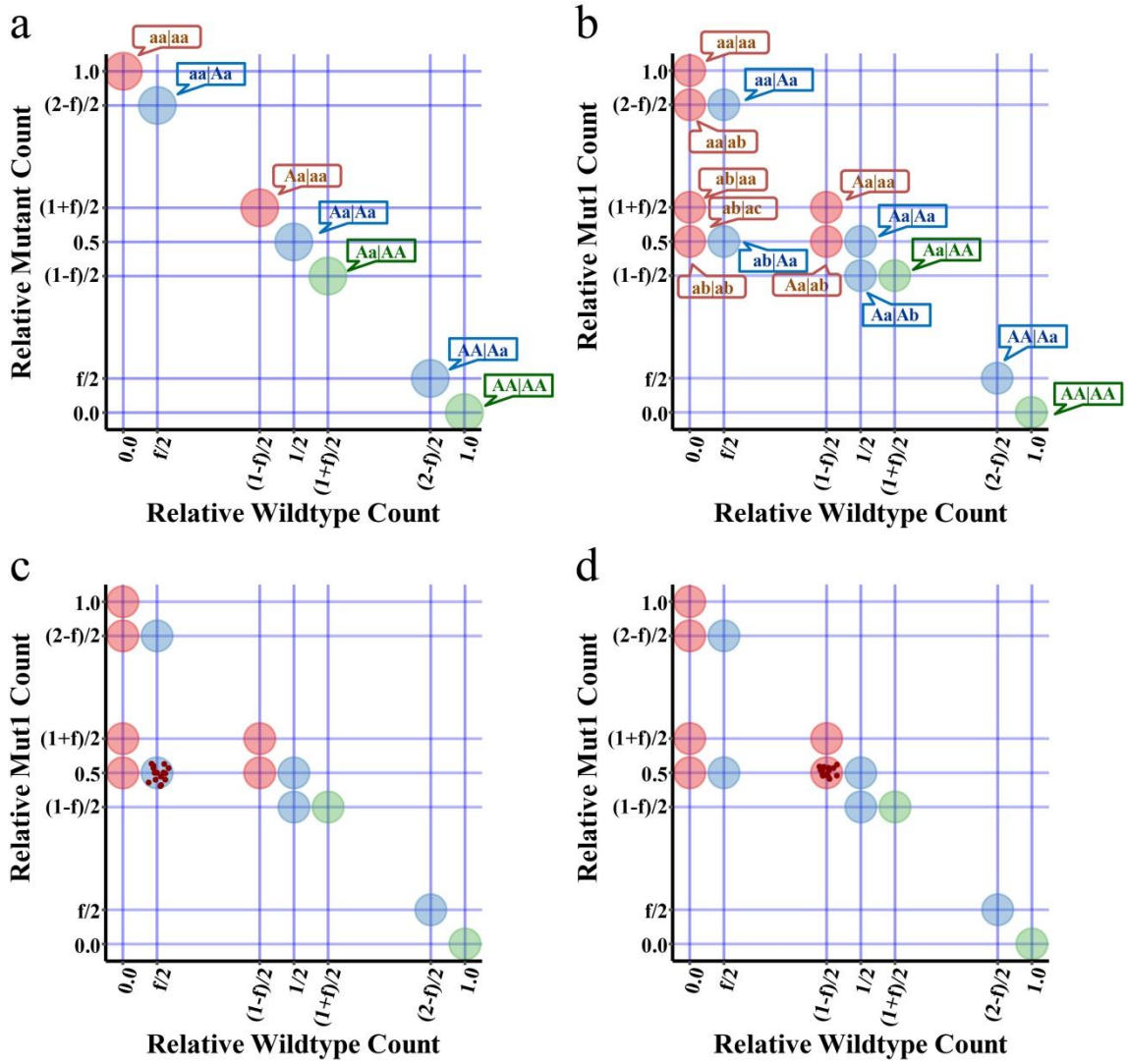


Fig. S19: detection of short genetic variation. a. expected possible positions for two-allele polymorphic sites on a normal disomy-disomy chromosome. b. expected all possible positions for polymorphic sites on a normal disomy-disomy chromosome. c. relative allelic count plot for a representative target site with library-level replicates. From the characteristic cluster position, the target site was estimated to be a heterozygous mutant-mutant for the mother and a heterozygous wildtype-mutant for the fetus. d. relative allelic count plot for a representative target site with library-level replicates. From the characteristic cluster position, the target site was estimated to be a heterozygous wildtype-mutant for the mother and a heterozygous mutant-mutant for the fetus, with the fetus carrying two different mutant alleles. A: wildtype allele. a-c: different mutant alleles.

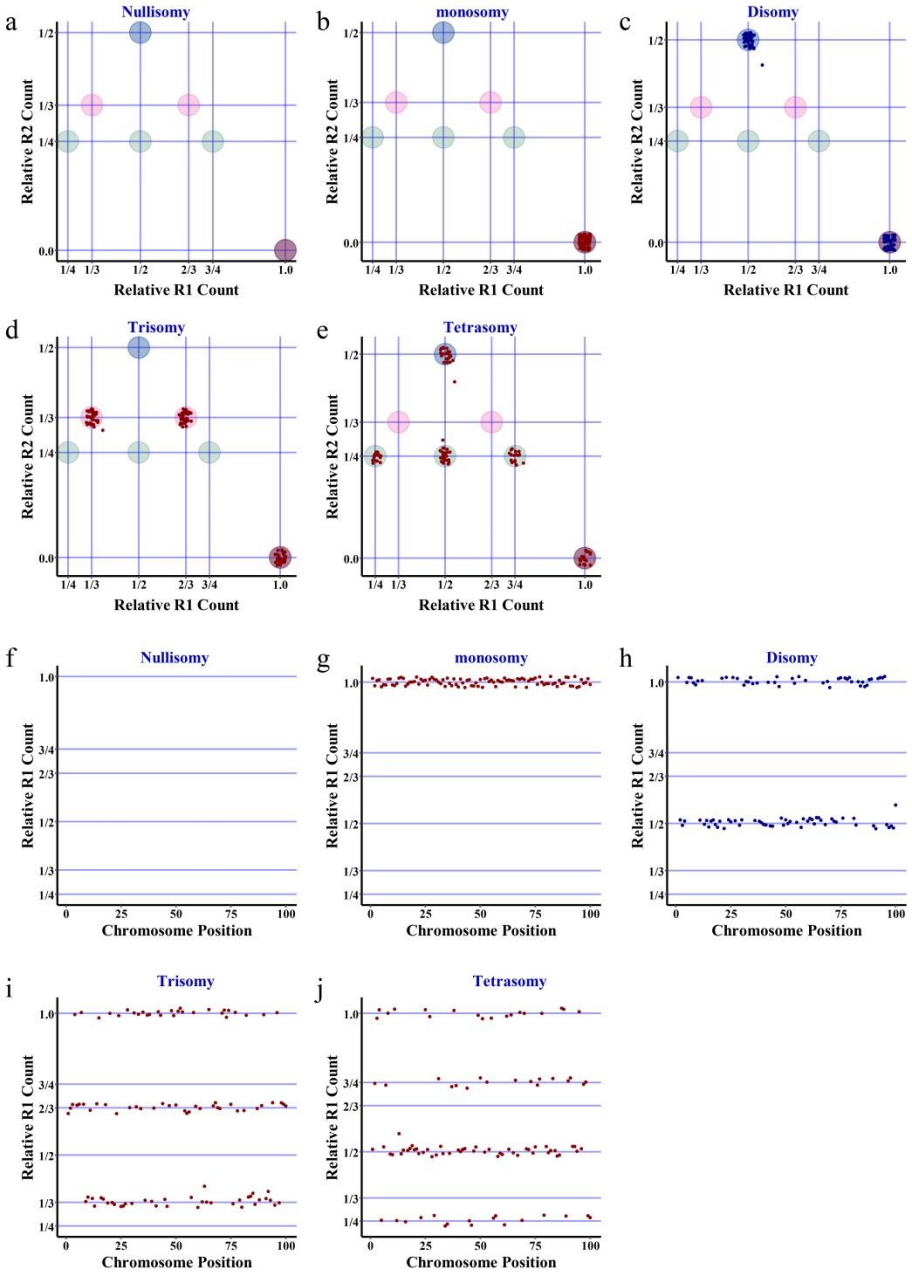


Fig. S19: detecting genetic aberrations for samples from non-pregnant individuals or preimplantation embryos. A panel of polymorphic sites on the target chromosome was simulated for a normal non-pregnant individual, and relative allelic counts for each polymorphic site were calculated. Then the relative R2 count was plotted against the relative R1 count (a-e) or the relative R1 count was plotted against its relative chromosomal position (f-j) for each amplicon. a, f: detection of nullisomy (or homozygous microdeletion). b, g: detection of monosomy (or heterozygous microdeletion). c, h: detection of the normal karyotype. d, i: detection of trisomy (or heterozygous microduplication). e, j: detection of tetrasomy (or homozygous microduplication).

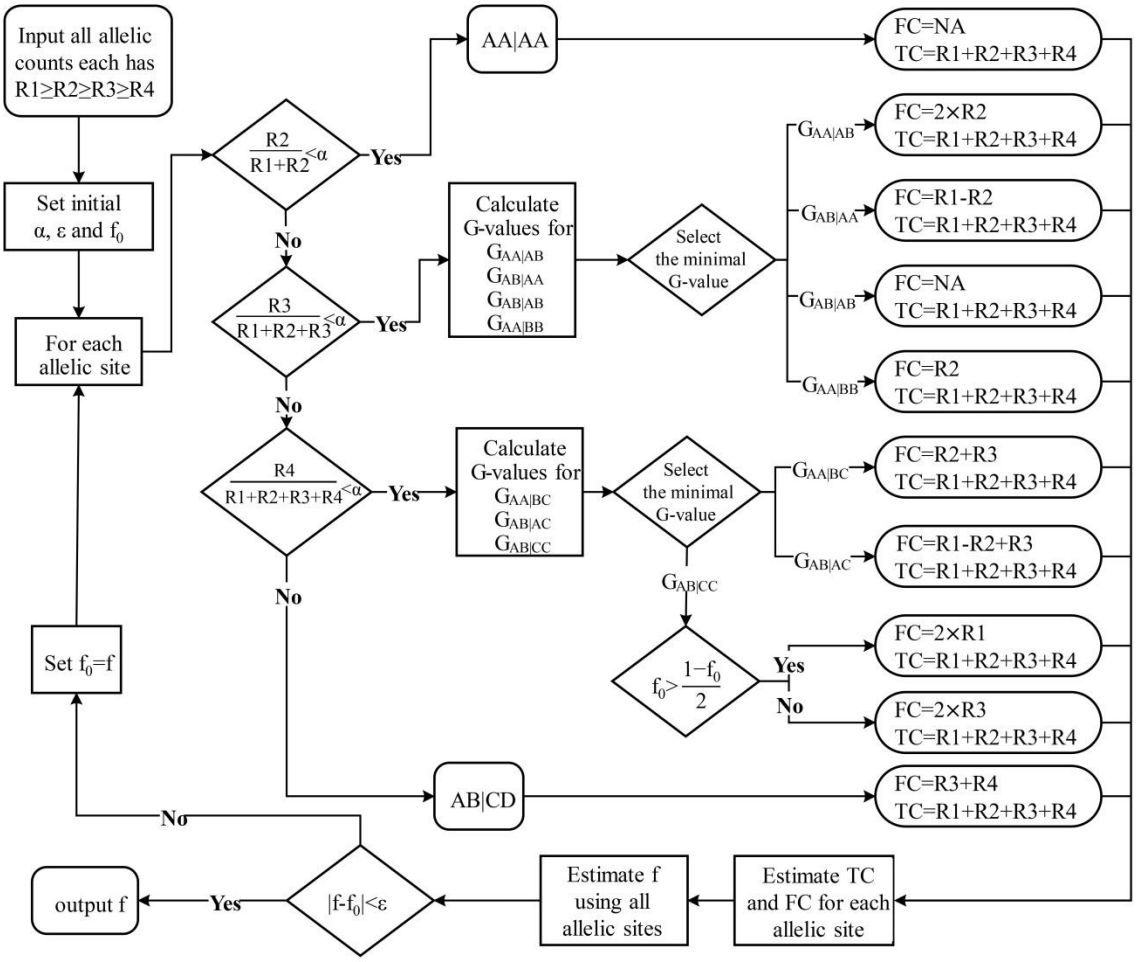


Fig. S21: estimating fetal fraction for a sample from a surrogate mother using allelic read counts. R1, R2, R3 and R4: allelic read counts in descending order; α: background threshold; ε: estimation precision; f_0: initial fetal fraction estimate; A-D: distinct alleles for each polymorphic site.

Table S1: All possible allelic read counts for a polymorphic site

Group	Genotype	Allelic Read Counts			R1 / Total Reads	Estimation	
		R1	R2	R3		Fetal Reads	Total Reads
I	AA|AA	$R_mR_mR_fR_f$	0	0	1	NA	R1
II	AA|AB	$R_mR_mR_f$	R_f	0	1-0.5*f	2.0*R2	R1+R2
III	AB|AA	$R_mR_fR_f$	R_m	0	0.5+0.5*f	R1-R2	R1+R2
IV	AB|AB	R_mR_f	R_mR_f	0	0.5	NA	R1+R2
V	AB|AC	R_mR_f	R_m	R_f	0.5	R1-R2+R3	R1+R2+R3

1. f: fetal fraction
2. R1, R2 and R3: read counts of each allele sorted in descending order
3. R_m and R_f: Reads mapped to maternal and fetal chromosomes, respectively.

Table S2: Genotype estimation for a two-allele site

Group	R1's Allele	R2's Allele	Expected Genotype	Relative Read Count	
				Wildtype	Mutant
I (AA|AA)	A		AA|AA	1	0
	a		aa|aa	0	1
II (AA|AB)	A	a	AA|Aa	(2-f)/2	f/2
	a	A	aa|Aa	f/2	(2-f)/2
III (AB|AA)	A	a	Aa|AA	(1+f)/2	(1-f)/2
	a	A	Aa|aa	(1-f)/2	(1+f)/2
IV (AB|AB)	A	a	Aa|Aa	1/2	1/2
	a	A	Aa|Aa	1/2	1/2

1. Group: estimated genotype groups by allelic read counts.
2. A and a: wildtype and mutant alleles for a polymorphic site

Table S3: Genotype estimation for a site with more than two alleles

Group	R1's Allele	R2's Allele	R3's Allele	Expected Genotype	Relative Read Count		
					Wildtype	Mutant 1	Mutant 2
I (AA|AA)	A			AA|AA	1	0	0
	a			aa|aa	0	1	0
II (AA|AB)	A	a		AA|Aa	(2-f)/2	f/2	0
	a	A		aa|Aa	f/2	(2-f)/2	0
	a	b		aa|ab	0	(2-f)/2	f/2
III (AB|AA)	A	a		Aa|AA	(1+f)/2	(1-f)/2	0
	a	A		Aa|aa	(1-f)/2	(1+f)/2	0
	a	b		ab|aa	0	(1+f)/2	(1-f)/2
IV (AB|AB)	A	a		Aa|Aa	1/2	1/2	0
	a	A		Aa|Aa	1/2	1/2	0
	a	b		ab|ab	0	1/2	1/2
V (AB|AC)	A	a	b	Aa|Ab	1/2	(1-f)/2	f/2
	a	A	b	Aa|ab	(1-f)/2	1/2	f/2
	a	b	A	ab|Aa	f/2	1/2	(1-f)/2
	a	b	c	ab|ac	0	1/2	(1-f)/2

1. Group: estimated genotype groups by allelic read counts. A-C: different alleles.
2. A and a-c: wildtype and mutant alleles for a polymorphic site