



Non-Invasive Prenatal Diagnosis of Monogenic Disorders Through Bayesian- and Haplotype-Based Prediction of Fetal Genotype

OPEN ACCESS

Edited by:

Zhonglin Jia,
Sichuan University, China

Reviewed by:

Xue-Ling Ou,
Sun Yat-sen University, China
Xiangdong Kong,
First Affiliated Hospital of Zhengzhou

University, China
Eftychia Dimitriadou,
University Hospitals Leuven, Belgium

*Correspondence:

Lijian Zhao
zhaolijian@genomics.cn
Xin Jin
jinxin@genomics.cn
Ya Gao
gaoya@genomics.cn
Aihua Yin
yinaihua@vip.126.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Human and Medical Genomics,
a section of the journal
Frontiers in Genetics

Received: 02 April 2022

Accepted: 13 June 2022

Published: 01 July 2022

Citation:

Li J, Lu J, Su F, Yang J, Ju J, Lin Y,
Xu J, Qi Y, Hou Y, Wu J, He W, Yang Z,
Wu Y, Tang Z, Huang Y, Zhang G,
Yang Y, Long Z, Cheng X, Liu P, Xia J,
Zhang Y, Wang Y, Chen F, Zhang J,
Zhao L, Jin X, Gao Y and Yin A (2022)
Non-Invasive Prenatal Diagnosis of
Monogenic Disorders Through
Bayesian- and Haplotype-Based
Prediction of Fetal Genotype.
Front. Genet. 13:911369.
doi: 10.3389/fgene.2022.911369

Jia Li^{1,2†}, Jiaqi Lu^{3†}, Fengxia Su^{4,5†}, Jiexia Yang^{6,7†}, Jia Ju^{4†}, Yu Lin⁴, Jinjin Xu⁴, Yiming Qi^{6,7}, Yaping Hou^{6,7}, Jing Wu^{6,7}, Wei He^{6,7}, Zhengtao Yang^{4,8}, Yujing Wu^{4,5}, Zhuangyuan Tang^{4,5}, Yingping Huang^{4,5}, Guohong Zhang^{4,5}, Ying Yang^{4,5}, Zhou Long⁴, Xiaofang Cheng⁴, Ping Liu⁴, Jun Xia⁴, Yanyan Zhang⁴, Yicong Wang⁴, Fang Chen⁴, Jianguo Zhang^{1,2}, Lijian Zhao^{1,2,9*}, Xin Jin^{4*}, Ya Gao^{4,5*} and Aihua Yin^{6,7*}

¹BGI Genomics, BGI-Shenzhen, Shenzhen, China, ²Hebei Industrial Technology Research Institute of Genomics in Maternal and Child Health, Shijiazhuang BGI Genomics, Shijiazhuang, China, ³Medical Genetics Centre, Guangdong Women and Children's Hospital, Guangzhou Medical University, Guangzhou, China, ⁴BGI-Shenzhen, Shenzhen, China, ⁵Shenzhen Engineering Laboratory for Birth Defects Screening, Shenzhen, China, ⁶Prenatal Diagnosis Centre, Guangdong Women and Children's Hospital, Guangzhou, China, ⁷Maternal and Children Metabolic-Genetic Key Laboratory, Guangdong Women and Children's Hospital, Guangzhou, China, ⁸College of Life Sciences, University of the Chinese Academy of Sciences, Beijing, China, ⁹College of Medical Technology, Hebei Medical University, Shijiazhuang, China

Background: Non-invasive prenatal diagnosis (NIPD) can identify monogenic diseases early during pregnancy with negligible risk to fetus or mother, but the haplotyping methods involved sometimes cannot infer parental inheritance at heterozygous maternal or paternal loci or at loci for which haplotype or genome phasing data are missing. This study was performed to establish a method that can effectively recover the whole fetal genome using maternal plasma cell-free DNA (cfDNA) and parental genomic DNA sequencing data, and validate the method's effectiveness in noninvasively detecting single nucleotide variations (SNVs), insertions and deletions (indels).

Methods: A Bayesian model was developed to determine fetal genotypes using the plasma cfDNA and parental genomic DNA from five couples of healthy pregnancy. The Bayesian model was further integrated with a haplotype-based method to improve the inference accuracy of fetal genome and prediction outcomes of fetal genotypes. Five pregnancies with high risks of monogenic diseases were used to validate the effectiveness of this haplotype-assisted Bayesian approach for noninvasively detecting indels and pathogenic SNVs in fetus.

Results: Analysis of healthy fetuses led to the following accuracies of prediction: maternal homozygous and paternal heterozygous loci, $96.2 \pm 5.8\%$; maternal heterozygous and paternal homozygous loci, $96.2 \pm 1.4\%$; and maternal heterozygous and paternal heterozygous loci, $87.2 \pm 4.7\%$. The respective accuracies of predicting insertions and deletions at these types of loci were $94.6 \pm 1.9\%$, $80.2 \pm 4.3\%$, and $79.3 \pm 3.3\%$. This approach detected pathogenic single nucleotide variations and deletions with an accuracy of 87.5% in five fetuses with monogenic diseases.

Conclusions: This approach was more accurate than methods based only on Bayesian inference. Our method may pave the way to accurate and reliable NIPD.

Keywords: non-invasive prenatal diagnosis, massively parallel sequencing, fetal genome, single nucleotide variations, monogenic disease

INTRODUCTION

Following the discovery of fetal cell-free DNA (cfDNA) in maternal plasma (Lo et al., 1997), next-generation sequencing technologies have enabled non-invasive prenatal screening for trisomies 13, 18, and 21; aneuploidies involving sex chromosomes (Chen et al., 2011; Agarwal et al., 2013; Benn et al., 2013; Mazloom et al., 2013; Samango-Sprouse et al., 2013; Hooks et al., 2014); and, more recently, rare autosomal aneuploidies and various sub-chromosomal aberrations (Snyder et al., 2015; Zhou et al., 2019). Maternal plasma cfDNA testing is now being applied to non-invasive prenatal diagnosis (NIPD) of monogenic diseases. So far, such testing has involved whole-exome sequencing of the cfDNA and analysis supplemented by parental haplotype information (Fan et al., 2012a; Lam et al., 2012; Yoo et al., 2015; Zhang et al., 2015). However, these methods can detect only autosomal dominant diseases and a few autosomal recessive diseases caused by known mutations. They cannot detect diseases that have not already been associated with mutation hotspots or that are caused by *de novo* variants (DNVs).

Another limitation of these methods is that they require proband genomic DNA to allow haplotype phasing, which may not be feasible if the proband passes away at the early life. To avoid this requirement, two groups developed methods to infer fetal genotype based on haplotyping of one or both parents. One method was able to detect only ~66%–70% of paternal-specific alleles and deduce only ~70% of paternally inherited haplotypes (Fan et al., 2012b), while the other method predicted heterozygous maternal and homozygous paternal (ABAA) loci with only 64.4% accuracy, and it was unable to predict many heterozygous maternal and paternal (ABAB) loci for lack of paternal haplotype information (Kitzman et al., 2012). Moreover, these methods cannot detect variants for which no haplotype or genome phasing information is available.

In the present study, we have established a Bayesian model that predicts fetal genotype based on whole-genome sequencing of cfDNA in maternal plasma and of single-tube long fragment reads (stLFRs) in paternal genomic DNA. This allows the reconstruction of parental haplotypes without the need of proband DNA, which in turn renders the fetal genotyping more accurate. We validate the effectiveness of our approach by non-invasively detecting single-nucleotide variants (SNVs), insertions and deletions (InDels) in five fetuses at risk of monogenic diseases.

METHODS AND MATERIALS

Study Design and Study Population

Five mothers with normal singleton pregnancies and their male partners were prospectively recruited into this study at

the Department of Fetal Medicine and Prenatal Diagnosis at Guangdong Women and Children's Hospital. All five pregnant women showed normal nuchal translucency (NT), and non-invasive prenatal screening results were negative for trisomies 13, 18, and 21. All women delivered healthy babies by vaginal delivery (**Supplementary Table S1**). Peripheral blood (5 ml) of each mother and father was sampled into an ethylenediamine tetraacetic acid-containing tube to provide information for the haplotype- and Bayesian-based method, the results of which were compared against umbilical cord blood (2 ml).

To validate our method for detecting pathogenic SNVs and indels for NIPD, another five mothers and their male partners whose fetuses were at risk of the following monogenic diseases were also prospectively recruited: tetrahydrobiopterin deficiency hyperphenylalafivemia, Duchenne/Becker muscular dystrophy, ocular albinism, muscular dystrophy polysaccharide glycosylation deficiency A11 and deafness. The five families were recruited because the parents were known to be carriers of disease alleles, or the fetuses were suspected of having monogenic diseases due to ultrasound abnormalities. All five pregnant women agreed to undergo amniocentesis for prenatal diagnosis, and maternal and paternal genomic DNA was Sanger-sequenced to confirm the presence of disease variants.

All families received a detailed explanation of the study and gave written informed consent before any samples were collected. The study strictly followed the Declaration of Helsinki and was approved by the Ethics Committee of the Guangdong Women and Children's Hospital (no. 201901091), as well as by the Institutional Review Board of the BGI (BGI-IRB 20002).

Preparation of cfDNA Libraries

Maternal blood was collected and within 8 h, it was centrifuged at 1,600 g for 10 min. Plasma was transferred to fresh microcentrifuge tubes and centrifuged at 16,000 g for 10 min to remove residual cells. From 600 μ l of the clarified plasma was extracted cfDNA using the MGIEasy Circulating DNA Isolation Kit (MGI, Shenzhen, China), which was used to construct a library with the MGIEasy Cell-free DNA Library Prep Kit (MGI, Shenzhen, China) based on a modified protocol (Xu et al., 2019). In brief, the extracted cfDNA was end-repaired, ligated with "A" tailing and then ligated with adapters. The ligated products were cleaned up and subjected to 10 cycles of PCR amplification. The PCR products were cleaned up, quantitated with a dsDNA Fluorescence Assay Kit (Invitrogen, United States), heat-denatured and incubated at 37°C to create ssDNA circles. These circles were subjected to rolling circle amplification to generate DNA nanoballs (Drmanac et al., 2010).

Preparation of Parental gDNA Libraries

High-molecular-weight parental genomic DNA was isolated from blood using a dialysis-based method (Wang et al., 2019) and prepared for stLFR sequencing, in which the same barcode sequence was added to subfragments of long DNA molecules to enable their second-generation sequencing (Wang et al., 2019). The resulting high-molecular-weight parental DNA (1.5 ng) was used to construct a library with the MGIEasy stLFR Library Prep Kit (MGI, Shenzhen, China). In brief, a hybridization sequence of 200–1,000 bp was added to the genomic DNA using transposons, and the resulting transposon-integrated DNA was allowed to adsorb onto beads. The transposons were ligated to barcode adapters, followed by other adaptors to allow multiplex sequencing. The ligated products were cleaned, subjected to five cycles of PCR amplification, purified and quantified using the Qubit® dsDNA HS Assay Kit (Invitrogen, United States).

Preparation for Umbilical Blood DNA Libraries

Umbilical blood DNA was extracted using an MGIEasy Magnetic Beads Genomic DNA Extraction Kit (MGI, Shenzhen, China), then used to prepare a library with the MGIEasy universal DNA Library Prep Set (MGI). Genomic DNA was fragmented with Segmentase (MGI, Shenzhen, China) to generate molecules 100–500 bp long, and fragments 280–320 bp were enriched using magnetic beads. The ends were filled in, then the base A was added to the 3' end to allow DNA fragments to be ligated to an adapter with base T at the 3' end. The DNA fragments were amplified by ligation-mediated PCR and purified to form the library.

Library Quality Control and Sequencing

Library control was checked using an Agilent DNA 1000 kit on a Bioanalyzer 2,100 platform (Agilent, United States) and quantified using a Qubit™ ssDNA Assay Kit (Invitrogen, United States). Then libraries were subjected to multiplex sequencing on a DNBSEQ platform (MGI, Shenzhen, China) according to a “paired-end 100 bp” strategy.

Read Mapping and Variant Calling

Raw reads were trimmed and filtered using SOAPnuke 2.1.1 (Chen et al., 2018). Reads were excluded if their N proportion was below 0.1, if > 50% of bases had a quality score < 12, or if there were > 2 mismatches with the adapter. The resulting “clean reads” were aligned to the human genome reference (hg19) using Burrows-Wheeler Aligner (BWA-mem) software (Li and Durbin, 2009), then duplicate reads were removed, InDels were realigned, and base quality scores were recalibrated using Sentieon genomics software2 (Kendig et al., 2019) and default parameters. Variants were called using Sentieon genomics software2. SNVs and InDels were detected using Sentieon DNAscope software (McKenna et al., 2010), which combines the GATK's HaplotypeCaller and a genotyping model based on machine learning.

Prediction of Fetal Genotype Using a Bayesian Model

FF was calculated by comparing the aligned sequence reads at maternal homozygous sites and fetal heterozygous sites using the formula $FF = 2p/(p + q) \times 100$, where p is the number of reads corresponding to fetal-specific alleles and q the number of reads shared between the mother and fetus. A Bayesian model was used to infer fetal genotype at each locus based on the FF and parental genotyping. The Bayesian model proceeded in two steps. First, the cumulative probability of the combination of each maternal and fetal genotype at each locus was calculated based on the read depth and FF at that locus. Second, the prior probabilities of the maternal and fetal genotype combinations were determined based on parental genotyping and Mendelian laws of inheritance. The Bayesian model generated 10 posterior probabilities, one for each possible combination of maternal and fetal genotypes. The predicted fetal genotype at each locus was the genotype with the highest posterior probability (Eq. 1)

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^n P(B|A_i)P(A_i)} \quad (1)$$

where $P(A_i)$ is the prior probability of the i th maternal and fetal genotype combination, calculated according to Mendelian laws; $P(B|A_i)$ is the cumulative probability of maternal and fetal genotype combination i based on read depth and FF at that locus; and n was 10 maternal and fetal genotype combinations.

For each locus, the probability of obtaining base j was calculated as follows: (Eq. 2):

$$P_j = B_{jF}/2 * C + B_{jM}/2 * (1 - C) \quad (2)$$

where B_{jF} , an integer between 0 and 2, indicates the number of bases j in $F1iF2i$; C represents the FF in the predetermined region; B_{jM} , also an integer between 0 and 2, indicates the number of bases j in $M1iM2i$; and j represents A, T, G or C. Based on the occurrence probability P_j of base j and the number of reads of A_j , the cumulative probability $P(F1iF2iM1iM2i)$ of each maternal and fetal genotype combination was determined as follows: (Eq. 3)

$$P(F1iF2iM1iM2i) = \sum P_j A_j / \sum A_j \quad (3)$$

where A_j stands for the read count for base j (A, T, G or C). Based on the $P(F1iF2iM1iM2i)$ of each maternal and fetal genotype combination, $P_{\text{final}}(F1iF2iM1iM2i)$ was computed according to the formula (Eq. 4)

$$P_{\text{final}}(F1iF2iM1iM2i) = \frac{P(F1iF2iM1iM2i)}{\sum (P(F1iF2iM1iM2i))} \quad (4)$$

The highest probability for the genotype combination was taken to be the final cumulative probability for that combination and was used in the Bayesian model. The predicted genotype at each locus was defined as the one with the highest posterior probability (patent PIDC3194001P).

In this way, the Bayesian model relied on parental variants and cfDNA preprocessing data as inputs, and it returned 10 posterior probabilities for 10 predicted fetal genotypes. Low-quality

variants were eliminated from the Bayesian analysis (**Supplementary Method S1**).

Prediction of Fetal Genotype Using a Haplotype-Based Method

We used a haplotype-based method based on sequential probability ratio testing (SPRT) and the “closest-variant” algorithm to predict fetal genotypes at AAAB and ABAB loci. First, we used longhap software (https://github.com/stLFR/stLFR_LongHap) to perform genome phasing based on the alignment and variant results of parental stLFR sequencing. Second, we deduced the inheritance of maternal haplotype at ABAA loci used a previous method (Lo et al., 2010). The SPRT was performed to determine whether the cumulative allele counts for SNVs along a haplotype block reached sufficient statistical confidence for Hap I or Hap II to be scored. SNVs for which statistical confidence was too low for a genotype call were considered “unclassified”. The maternally contributed alleles for these unclassified variants were inferred using the closest-variant algorithm 1, which predicted maternal or paternal inheritance. based on the inferred inheritance of the nearest variant within 200 kb in the same haplotype block. If the upstream and downstream variants showed different inherited haplotypes within a 200-kb region, these unclassified variants were not analyzed (**Supplementary Figure S1**). The closest-variant algorithm 1 was defined to predict maternal or paternal inheritance based on the inferred inheritance of the nearest variant within 200 kb in the same haplotype block. We defined the closest variant algorithm 2 to infer the paternally/maternally contributed allele for the variant using the inferred inheritance of the nearest variant within 500 Kb region of the same haplotype block (**Supplementary Figure S1**).

We defined the closest variant algorithm 2 to infer the paternally/maternally contributed allele for the variant using the inferred inheritance of the nearest variant within 500 Kb region of the same haplotype block (**Supplementary Figure S1**). In the case of InDels, we used the closest-variant algorithm 2 to infer paternal inheritance at AAAB and ABAB loci, and SPRT to predict maternal inheritance. The closest-variant algorithm 2 was also used to infer the maternally contributed alleles of unclassified variants. Low-confidence variants were filtered out in the SPRT (**Supplementary Method S2**).

Combination of Bayesian- and Haplotype-Based Prediction of Fetal Genotype

We used the Bayesian model to infer paternally inherited alleles at AAAB loci, while we used the haplotype-based method to infer maternally inherited alleles at ABAA loci. In the case of ABAB loci, we first used the closest-variant algorithm 1 to determine the paternally inherited alleles based on the inferred inheritance at the closest AAAB locus determined by the Bayesian model within 200 kb in the same haplotype block, after which we conducted

SPRT to predict maternally inherited alleles. The closest-variant algorithm 1 was used to determine maternal alleles of unclassified variants. Finally, we used the Bayesian model to predict fetal genotype at the remaining ABAA and ABAB loci for which the haplotype-based method did not predict genotype (**Figure 2**).

RESULTS

Bayesian Model for Inferring Fetal Genotype

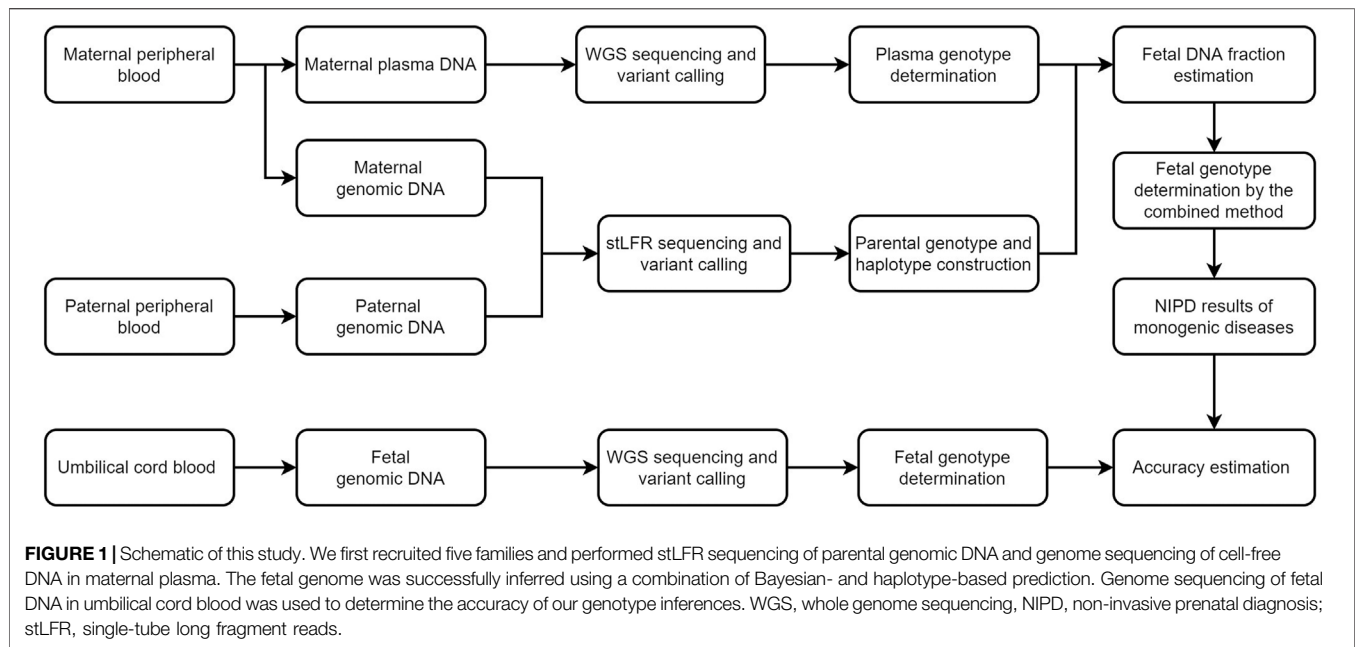
Plasma cfDNA and genomic DNA from five healthy pregnant women and their husbands were sequenced. The women were aged 30.72 years (range, 28.5–32.6 years) bearing fetuses at a mean gestational age of 24.2 weeks (range, 13–33 weeks) (**Supplementary Table S1**). The cfDNA was sequenced at a depth of 100X (range, 112.03–256.12X); the genomic DNA, to a depth >30X (range, 28.77–68.39X; **Supplementary Table S2, Figure 1**). Umbilical cord blood DNA was sequenced at a depth of 48X (range, 41.85–52.34X; **Supplementary Table S2**). The estimated FF had a mean of 13% (range, 4%–27%; **Supplementary Table S1**).

A Bayesian model to infer fetal SNVs showed the greatest accuracy with fetus JK-16, who had the highest FF (27%), and lowest accuracy with JK-53, who had the lowest FF (4%). These results indicate the strong influence of FF on fetal genotype inference. Increasing the cfDNA sequencing depth from 128 to 256.12X increased the accuracy of JK-53 genotyping (**Table 1**). The genotyping accuracy across the five families was $96.2 \pm 5.8\%$ at homozygous maternal and heterozygous paternal loci (AAAB), $74.6 \pm 9.5\%$ at ABAA loci, and $64.3 \pm 11.9\%$ at ABAB loci (**Table 1; Figure 2**).

Improving the Accuracy of Fetal Genotype Inference Using Haplotyping

Given the Bayesian model’s poor performance at predicting fetal genotypes at ABAA and ABAB loci, we genotyped complex haploid subsets of maternal and paternal genomic DNA at these loci while preserving long-range contiguity (**Figure 2**). We directly phased over 99% of ABAA loci into long haplotype blocks, giving an average N50 of 18.72 Mb, and over 99% of ABAB loci into long haplotype blocks, giving an average N50 of 13.57 Mb (**Supplementary Table S2**). This haplotype-based method successfully classified 90%–97% of maternally inherited SNVs at ABAA loci and correctly predicted 98%–99% of SNVs (**Table 1**). Nevertheless, this haplotype-based method was unable to infer genotype at 3%–10% of ABAA loci, so we inferred these gaps using the Bayesian model. This combination of haplotype-based and Bayesian-based prediction (hereinafter referred to the combined method) gave SNV genotyping accuracies of 94%–98% at ABAA loci (**Table 1; Figure 3A**). The haplotype-based method successfully classified 74%–92% of maternally inherited SNVs at ABAB loci and correctly predicted 93%–98% of SNVs. Adding Bayesian inference to fill in gaps led to accuracies of 82%–95% (**Table 1; Figure 3B**).

We also utilized the haplotype-based method to infer paternal and maternal inheritances for fetal InDels, which was successful for 95% of all InDels (range, 93.6%–97%) across the fetuses.



Accuracy was highest at ABAA loci (94.6%), followed by AAAB loci (80.2%) and ABAB loci (79.3%) (**Supplementary Table S3, Supplementary Figure S2**).

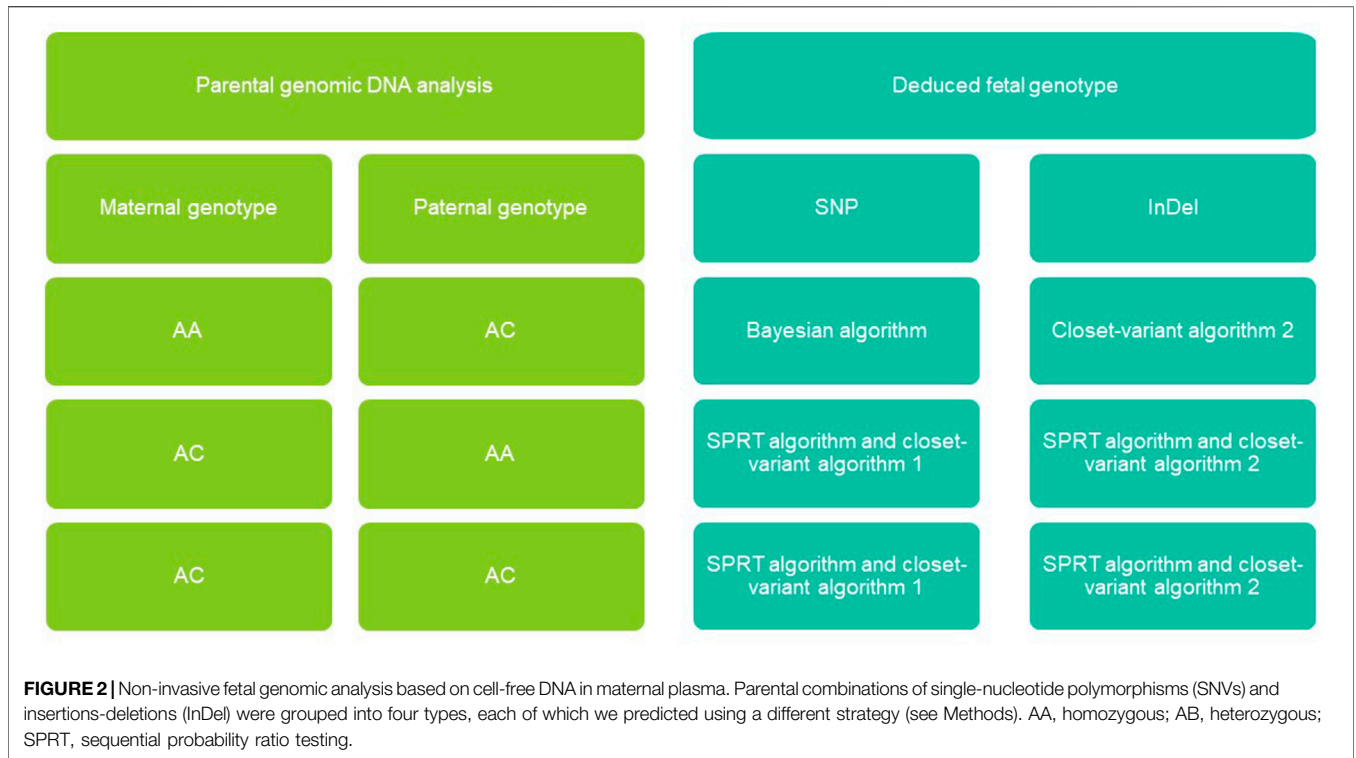
Combining Haplotype- and Bayesian-Based Prediction for NIPD of Monogenic Diseases

We sequenced the cfDNA from the plasma of five pregnant women with a mean age of 29 years (range, 25–34 years) at a mean gestational age of 13 weeks (range, 11–18 weeks) whose

fetuses were at risk for monogenic diseases. Mean FF was 0.15 (range, 0.08–0.20; **Table 2**). The cfDNA was sequenced at a depth of 121X (**Supplementary Table S4**). In parallel, stLFR sequencing of parental genomic DNA was performed at a depth of 24X. The combination of haplotype- and Bayesian-based prediction identified 8 pathogenic variants in five genes in the plasma cfDNA from all five mothers (**Table 2, Supplementary Table S5**). Our method correctly identified 6 heterozygous carriers of monogenic disease variants, including tetrahydrobiopterin deficiency hyperphenylalafivemia, Duchenne/Becker muscular

TABLE 1 | Performance metrics for inferring fetal SNPs in 5 healthy families.

Family	Heterozygous	The bayesian model		The haplotype-based method		The combined method	
		Accuracy (%)	Number of true predictions/ number of loci	Accuracy (%)	Number of true predictions/ number of loci	Accuracy (%)	Number of true predictions/ number of loci
JK-7	AAAB loci	96.8	631788/652391				
	ABAA loci	72.6	909990/1253908	98.8	1171962/1186610	96.1	1209126/1253908
	ABAB loci	61.4	429479/698962	96.4	572296/593618	89.8	627424/698962
JK-16	AAAB loci	97.7	576306/590073				
	ABAA loci	91.8	1014224/1104405	99.2	1061980/1070070	98	1082868/1104405
	ABAB loci	85.9	540677/629324	97.5	562304/576574	94.9	597082/629324
JK-18	AAAB loci	96.4	463476/480708				
	ABAA loci	75.7	901292/1190790	99	1112107/1123516	96.7	1151107/1190790
	ABAB loci	65.9	439900/667062	92.7	488305/526841	86.5	576938/667062
JK-28	AAAB loci	95.2	313500/329369				
	ABAA loci	69.5	440760/633774	98.5	581675/590448	96	608483/633774
	ABAB loci	57.3	222910/388712	93.1	268695/288587	83.2	323556/388712
JK-53 (128X)	AAAB loci	95.6	540046/564870				
	ABAA loci	59.6	667583/1120831	97.3	926919/952660	92.0	998262/1085313
	ABAB loci	49.4	325416/658784	89.1	363047/407606	74.7	481240/644645
JK-53 (256.12X)	AAAB loci	95	590368/621392				
	ABAA loci	63.5	706502/1112899	97.7	979699/1002260	93.6	1041875/1112899
	ABAB loci	51	341350/668854	93.9	462865/492889	81.8	547115/668854



dystrophy, ocular albinism, muscular dystrophy polysaccharide glycosylation deficiency A11 and deafness and 1 wildtype variant in a fetus at risk of Tetrahydrobiopterin deficiency hyperphenylalaninemia. The method also correctly predicted the heterozygous deletion c.8371delC (CDH23 in NM_022124) in a fetus at risk of

deafness. The inferred fetal variants were validated by Sanger sequencing of DNA from umbilical cord blood, which revealed only one incorrect inference (**Table 2, Supplementary Table S5**). Therefore, in this sample of five fetuses, our method was able to non-invasively determine pathogenic variants with an accuracy of 87.5%.

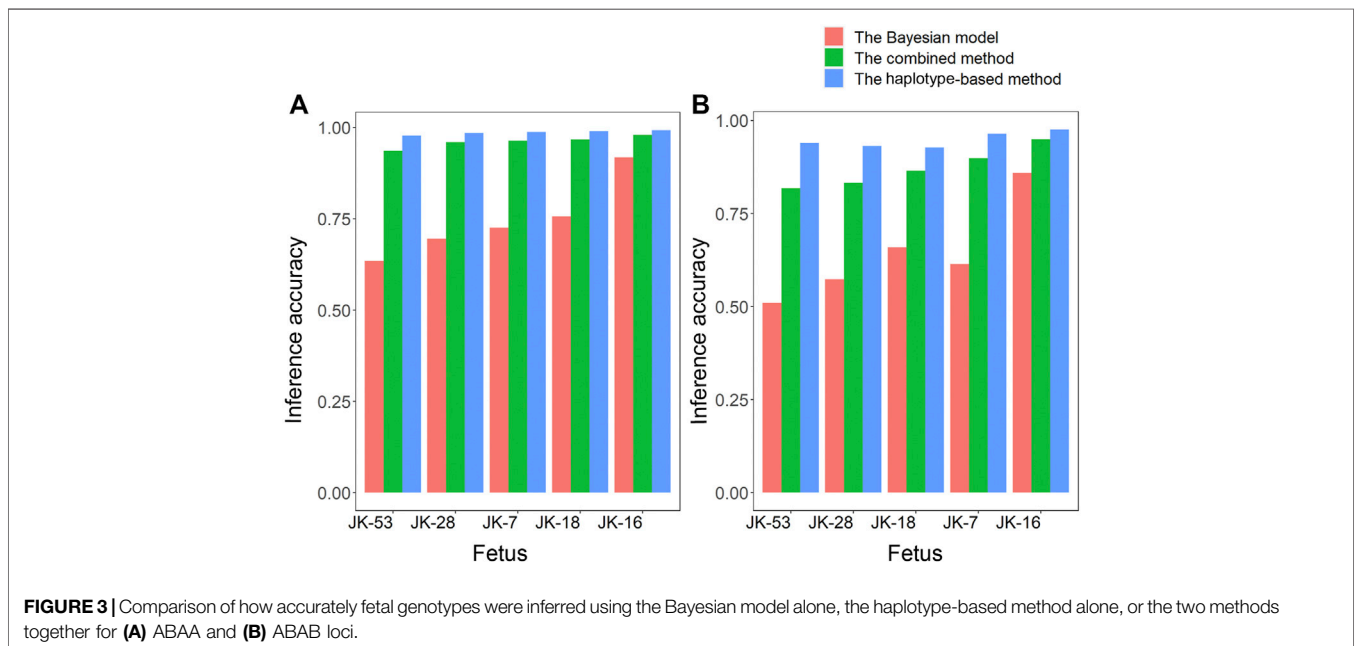


TABLE 2 | Summary of non-invasive prenatal diagnosis in 5 families with monogenic diseases by the combined model.

Family ID	Age	Gestational week	FF (%)	Monogenic diseases	Maternal genotype	Paternal genotype	Inferred fetal genotype (bayesian model prediction/haplotype-based prediction)
SFY-10	25	12	16	Tetrahydrobiopterin deficiency hyperphenylalaninemia	PTS(NM_000317) heterozygous c.73C > G	PTS(NM_000317) heterozygous c.155A > G	C/C (C/C, NA) A/G (A/G, A/G)
SFY-15	34	11	16	Duchenne/Becker Muscular Dystrophy	DMD (NM_004006) heterozygous c.187–2A > T	Wild-type	A/T (A/T, NA)
SFY-32	29	11	8	Deafness	CDH23(NM_022124) heterozygous c.8371delC	CDH23(NM_022124) heterozygous c.1606C > T	AC/A (NA, AC/A) C/T (C/T, NA)
SFY-05	27	18	20	Muscular dystrophy polysaccharide glycosylation deficiency A11	B3GALNT2(NM_152490) heterozygous c.181C > T	B3GALNT2(NM_152490) heterozygous c.261–2A > G	C/C (C/C, not in block) ^a A/G (A/G, NA)
SFY-18	30	13	14	Ocular albinism	GPR143 heterozygous c.885 + 748G > A	Wild-type	G/A (G/A, G/A)

^aIndicates incorrect prediction.

DISCUSSION

In this study, we developed a Bayesian model for non-invasively inferring fetal genotypes based on sequencing of cfDNA in maternal plasma and of parental genomic DNA. The model accurately predicted fetal genotype at AAAB loci but poorly at ABAA and ABAB loci. By combining this approach with haplotype information, we accurately predicted SNVs and InDels at AAAB, ABAA and ABAB loci with high prediction accuracy despite a relatively low FF. We demonstrated the potential of our combined method for NIPD of monogenic diseases.

Over the past decade, several haplotype-based strategies have been reported for inferring fetal genotypes based on deep sequencing of maternal plasma (Lo et al., 2010; Fan et al., 2012b; Kitzman et al., 2012; Chen et al., 2013; Chan et al., 2016). In these strategies, the parental haplotype is determined directly (Kitzman et al., 2011; Lam et al., 2012) or derived from analysis of pedigrees (Lo et al., 2010; Chen et al., 2013; New et al., 2014) or founder haplotypes in selected populations (Zeevi et al., 2015). Haplotype blocks in these strategies average in size from 300 kb to >1 Mb, which restricts the resolution at which maternal inheritance of the fetus can be inferred (Lo et al., 2010; Fan et al., 2012b; Kitzman et al., 2012). The lack of complete haplotype information or genome phasing information in these strategies means that only ~70% of paternally inherited haplotypes or ABAB loci can be analyzed (Fan et al., 2012b; Kitzman et al., 2012). A Bayesian method has been reported that can predict SNVs and InDels independently of the inheritance model and parental origin, but it cannot detect DNVs, multi-allelic loci or X-linked inheritance (Rabinowitz et al., 2019).

Our haplotype-based method was able to infer genotypes at AAAB, ABAA and ABAB loci with high accuracy, yet it could not do so for 3%–10% of ABAA loci or 8%–26% of ABAB loci. Therefore, we used a Bayesian model to predict the loci missed by the haplotype-based method. This combined approach

allowed the accurate prediction of SNVs and InDels at all fetal loci at single-base resolution. Our method appears to be able to infer fetal genotype with much higher resolution than previously reported methods. For example, one previous method predicted only a fraction of ABAA loci, whereas it was unable to analyze ABAB loci (Chan et al., 2016). Another method predicted SNVs at ABAA loci with an accuracy of only 64.4%, but it was able to analyze SNVs at only some ABAB loci for lack of paternal haplotype information (Fan et al., 2012b). In contrast, our method correctly predicted SNVs at AAAB, ABAA and ABAB loci with accuracies of 82%–95%. In addition, our method accurately predicted SNVs and InDels. For example, our method predicted Indels at AAAB, ABAA and ABAB loci with accuracies of 79%–95%. In fact, during analysis of five fetuses at risk of monogenic disease, our method detected five disease-causing mutations. Thus, our method appears to be the only one reported so far that can comprehensively predict SNVs and InDels. Moreover, our method delivered accurate predictions at FFs as low as 4%, much lower than in previously published methods (Kitzman et al., 2012; Rabinowitz et al., 2019), indicating the potential for NIPD early in pregnancy.

By using stLFR technology we were able to determine parental haplotypes without the need for proband DNA and with a much shallower sequencing depth than a previously published method (Chan et al., 2016). Our approach may become accessible to more institutions as genome-wide direct phasing becomes less expensive and technically demanding (Che et al., 2020). At the same time, our method needs to be improved to increase its clinical feasibility, such as increasing the accuracy of predicting SNVs at ABAB loci and InDels at AAAB and ABAB loci. The Bayesian model that we applied here calculates the likelihood of the fetal genotype using the maternal genotype and FF. Adding other features to the model may improve its ability to discriminate fetal and maternal reads; such features may include fragment size (Rabinowitz et al., 2019) and clusters of preferred ending positions of fetal

fragments (Chan et al., 2016). Another approach to improve inference accuracy may be to apply scalable FF amplification technology (Welker et al., 2020).

CONCLUSION

We have established a haplotype- and Bayesian-based method that can accurately predict fetal genotype at single-base resolution. Our method may be useful for accurately recovering fetal genomes and for NIPD of monogenic diseases caused by SNVs or InDels.

DATA AVAILABILITY STATEMENT

The data reported in this study are available from the CNGB Sequence Archive in the CNGBdb database under accession number CNP0001437.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The study strictly followed the Declaration of Helsinki and was approved by the Ethics Committee of the Guangdong Women and Children's Hospital (no. 201901091), as well as by the Institutional Review Board of the BGI (BGI-IRB 20002). The patients/participants provided their written informed consent to participate in this study.

REFERENCES

- Agarwal, A., Sayres, L. C., Cho, M. K., Cook-Deegan, R., and Chandrasekharan, S. (2013). Commercial Landscape of Noninvasive Prenatal Testing in the United States. *Prenat. Diagn.* 33, 521–531. doi:10.1002/pd.4101
- Benn, P., Cuckle, H., and Pergament, E. (2013). Non-invasive Prenatal Testing for Aneuploidy: Current Status and Future Prospects. *Ultrasound Obstet. Gynecol.* 42, 15–33. doi:10.1002/uog.12513
- Chan, K. C. A., Jiang, P., Sun, K., Cheng, Y. K. Y., Tong, Y. K., Cheng, S. H., et al. (2016). Second Generation Noninvasive Fetal Genome Analysis Reveals De Novo Mutations, Single-Base Parental Inheritance, and Preferred DNA Ends. *Proc. Natl. Acad. Sci. U.S.A.* 113, E8159–E8168. doi:10.1073/pnas.1615800113
- Che, H., Villella, D., Dimitriadou, E., Melotte, C., Brison, N., Neofytou, M., et al. (2020). Noninvasive Prenatal Diagnosis by Genome-wide Haplotyping of Cell-free Plasma DNA. *Genet. Med.* 22, 962–973. doi:10.1038/s41436-019-0748-y
- Chen, E. Z., Chiu, R. W. K., Sun, H., Akolekar, R., Chan, K. C. A., Leung, T. Y., et al. (2011). Noninvasive Prenatal Diagnosis of Fetal Trisomy 18 and Trisomy 13 by Maternal Plasma Dna Sequencing. *PLoS One* 6, e21791–7. doi:10.1371/journal.pone.0021791
- Chen, F. Z., You, L. J., Yang, F., Wang, L. N., Guo, X. Q., Gao, F., et al. (2020). CNGBdb: China National GeneBank DataBase. *Yi Chuan* 42, 799–809. doi:10.16288/j.ycz.20-080
- Chen, S., Ge, H., Wang, X., Pan, X., Yao, X., Li, X., et al. (2013). Haplotype-assisted Accurate Non-invasive Fetal Whole Genome Recovery through Maternal Plasma Sequencing. *Genome Med.* 5, 18. doi:10.1186/gm422
- Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S., et al. (2018). SOAPnuke: a MapReduce Acceleration-Supported Software for Integrated Quality Control and Preprocessing of High-Throughput Sequencing Data. *Gigascience* 7, 1–6. doi:10.1093/gigascience/gix120

AUTHOR CONTRIBUTIONS

Conceptualization: AY, YG, LZ Methodology: JaL, FS, JJ, YL, JjX, YY, and ZY Investigation: JqL, JY, YQ, YaH, JW, WH, YW, ZT, YiH, GZ, ZL, XC, PL, JX, YZ, YW, FC, and JZ Visualization: JaL, FS, JJ Funding acquisition: AY, YG, LZ, JX Project administration: FS, JqL Supervision: AY, YG, LZ, and XJ Writing-original draft: JaL, FS, JJ Writing-review and editing: YG, AY.

FUNDING

This work was funded by the National Nature Science Foundation of China (grant 81771598 to AY), National Key Research, National Key Research and Development Program of China (grant 2016YFC1000703 to AY), Guangzhou Science and Technology Planning Project (grant 202103000047 to AY), Medical Scientific Research Foundation of Guangdong Province of China (grant B2022082 to YQ) and Shenzhen Municipal Government of China (grant JCYJ20180703093402288 to YG).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.911369/full#supplementary-material>

- Drmanac, R., Sparks, A. B., Callow, M. J., Halpern, A. L., Burns, N. L., Kermani, B. G., et al. (2010). Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays. *Science* 327, 78–81. doi:10.1126/science.1181498
- Fan, H. C., Gu, W., Wang, J., Blumenfeld, Y. J., El-Sayed, Y. Y., and Quake, S. R. (2012a). Erratum: Non-invasive Prenatal Measurement of the Fetal Genome. *Nature* 489, 326. doi:10.1038/nature11423
- Fan, H. C., Gu, W., Wang, J., Blumenfeld, Y. J., El-Sayed, Y. Y., and Quake, S. R. (2012b). Non-invasive Prenatal Measurement of the Fetal Genome. *Nature* 487, 320–324. doi:10.1038/nature11251
- Guo, X., Chen, F., Gao, F., Li, L., Liu, K., You, L., et al. (2020). CNSA: a Data Repository for Archiving Omics Data. *Database (Oxford)* 2020, 1–6. doi:10.1093/database/baaa055
- Hooks, J., Wolfberg, A. J., Wang, E. T., Struble, C. A., Zahn, J., Juneau, K., et al. (2014). Non-invasive Risk Assessment of Fetal Sex Chromosome Aneuploidy through Directed Analysis and Incorporation of Fetal Fraction. *Prenat. Diagn.* 34, 496–499. doi:10.1002/pd.4338
- Kendig, K. I., Baheti, S., Bockol, M. A., Drucker, T. M., Hart, S. N., Heldenbrand, J. R., et al. (2019). Sentieon DNaseq Variant Calling Workflow Demonstrates Strong Computational Performance and Accuracy. *Front. Genet.* 10, 736. doi:10.3389/fgene.2019.00736
- Kitzman, J. O., Mackenzie, A. P., Adey, A., Hiatt, J. B., Patwardhan, R. P., Sudmant, P. H., et al. (2011). Haplotype-resolved Genome Sequencing of a Gujarati Indian Individual. *Nat. Biotechnol.* 29, 59–63. doi:10.1038/nbt.1740
- Kitzman, J. O., Snyder, M. W., Ventura, M., Lewis, A. P., Qiu, R., Simmons, L. E., et al. (2012). Noninvasive Whole-Genome Sequencing of a Human Fetus. *Sci. Transl. Med.* 4, doi:10.1126/scitranslmed.3004323
- Lam, K.-W. G., Jiang, P., Liao, G. J. W., Chan, K. C. A., Leung, T. Y., Chiu, R. W. K., et al. (2012). Noninvasive Prenatal Diagnosis of Monogenic Diseases by Targeted Massively Parallel Sequencing of Maternal Plasma: Application to β -Thalassemia. *Clin. Chem.* 58, 1467–1475. doi:10.1373/clinchem.2012.189589

- Li, H., and Durbin, R. (2009). Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* 25, 1754–1760. doi:10.1093/bioinformatics/btp324
- Lo, Y. M. D., Chan, K. C. A., Sun, H., Chen, E. Z., Jiang, P., Lun, F. M. F., et al. (2010). 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
- Lo, Y. M. D., Corbetta, N., Chamberlain, P. F., Rai, V., Sargent, I. L., Redman, C. W., et al. (1997). Presence of Fetal DNA in Maternal Plasma and Serum. *Lancet* 350, 485–487. doi:10.1016/S0140-6736(97)02174-0
- Mazloom, A. R., Džakula, Ž., Oeth, P., Wang, H., Jensen, T., Tynan, J., et al. (2013). Noninvasive Prenatal Detection of Sex Chromosomal Aneuploidies by Sequencing Circulating Cell-free DNA from Maternal Plasma. *Prenat. Diagn.* 33, 591–597. doi:10.1002/pd.4127
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytksy, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data. *Genome Res.* 20, 1297–1303. doi:10.1101/gr.107524.110
- New, M. I., Tong, Y. K., Yuen, T., Jiang, P., Pina, C., Chan, K. C. A., et al. (2014). Noninvasive Prenatal Diagnosis of Congenital Adrenal Hyperplasia Using Cell-free Fetal DNA in Maternal Plasma. *J. Clin. Endocrinol. Metab.* 99, E1022–E1030. doi:10.1210/jc.2014-1118
- Rabinowitz, T., Polsky, A., Golan, D., Danilevsky, A., Shapira, G., Raff, C., et al. (2019). Bayesian-based Noninvasive Prenatal Diagnosis of Single-Gene Disorders. *Genome Res.* 29, 428–438. doi:10.1101/gr.235796.118
- Samango-Sprouse, C., Banjevic, M., Ryan, A., Sigurjonsson, S., Zimmermann, B., Hill, M., et al. (2013). SNP-based Non-invasive Prenatal Testing Detects Sex Chromosome Aneuploidies with High Accuracy. *Prenat. Diagn.* 33, 643–649. doi:10.1002/pd.4159
- Snyder, M. W., Simmons, L. E., Kitzman, J. O., Coe, B. P., Henson, J. M., Daza, R. M., et al. (2015). Copy-Number Variation and False Positive Prenatal Aneuploidy Screening Results. *N. Engl. J. Med.* 372, 1639–1645. doi:10.1056/NEJMoa1408408
- Wang, O., Chin, R., Cheng, X., Wu, M. K. Y., Mao, Q., Tang, J., et al. (2019). Efficient and Unique Cobarcoding of Second-Generation Sequencing Reads from Long DNA Molecules Enabling Cost-Effective and Accurate Sequencing, Haplotyping, and De Novo Assembly. *Genome Res.* 29, 798–808. doi:10.1101/gr.245126.118
- Welker, N. C., Lee, A. K., Kjolby, R. A. S., Wan, H. Y., Theilmann, M. R., Jeon, D., et al. (2021). High-throughput Fetal Fraction Amplification Increases Analytical Performance of Noninvasive Prenatal Screening. *Genet. Med.* 23, 443–450. doi:10.1038/s41436-020-01009-5
- Xu, Y., Lin, Z., Tang, C., Tang, Y., Cai, Y., Zhong, H., et al. (2019). A New Massively Parallel Nanoball Sequencing Platform for Whole Exome Research. *BMC Bioinforma.* 20, 153. doi:10.1186/s12859-019-2751-3
- Yoo, S.-K., Lim, B. C., Byeun, J., Hwang, H., Kim, K. J., Hwang, Y. S., et al. (2015). Noninvasive Prenatal Diagnosis of Duchenne Muscular Dystrophy: Comprehensive Genetic Diagnosis in Carrier, Proband, and Fetus. *Clin. Chem.* 61, 829–837. doi:10.1373/clinchem.2014.236380
- Zeevi, D. A., Altarescu, G., Weinberg-Shukron, A., Zahdeh, F., Dinur, T., Chicco, G., et al. (2015). Proof-of-principle Rapid Noninvasive Prenatal Diagnosis of Autosomal Recessive Founder Mutations. *J. Clin. Invest.* 125, 3757–3765. doi:10.1172/JCI79322
- Zhang, X.-L., Qiu, X.-B., Yuan, F., Wang, J., Zhao, C.-M., Li, R.-G., et al. (2015). TBX5 Loss-Of-Function Mutation Contributes to Familial Dilated Cardiomyopathy. *Biochem. Biophys. Res. Commun.* 459, 166–171. doi:10.1016/j.bbrc.2015.02.094
- Zhou, Q., Zhu, Z.-P., Zhang, B., Yu, B., Cai, Z.-M., and Yuan, P. (2019). Clinical Features and Pregnancy Outcomes of Women with Abnormal Cell-free Fetal DNA Test Results. *Ann. Transl. Med.* 7, 317. doi:10.21037/atm.2019.06.57

Conflict of Interest: Authors JL, JZ, and LZ were employed by BGI Genomics, BGI-Shenzhen. FS, JJ, YL, JIX, ZY, YW, ZT, YH, GZ, YY, ZL, XC, PL, JuX, YZ, YW, FC, XJ, and YG BGI-Shenzhen.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Li, Lu, Su, Yang, Ju, Lin, Xu, Qi, Hou, Wu, He, Yang, Wu, Tang, Huang, Zhang, Yang, Long, Cheng, Liu, Xia, Zhang, Wang, Chen, Zhang, Zhao, Jin, Gao and Yin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.