




Noninvasive prenatal diagnosis by genome-wide haplotyping of cell-free plasma DNA

Huiwen Che, MS¹, Darine Villela, PhD^{1,2}, Eftychia Dimitriadou, PhD¹, Cindy Melotte, PhD¹, Nathalie Brison, PhD¹, Maria Neofytou, PhD¹, Kris Van Den Bogaert, PhD¹, Olga Tsuiko, PhD¹, Koen Devriendt, MD, PhD¹, Eric Legius, MD, PhD¹, Masoud Zamani Esteki, PhD^{1,3,4}, Thierry Voet, PhD¹ and Joris Robert Vermeesch, PhD¹ 

Purpose: Whereas noninvasive prenatal screening for aneuploidies is widely implemented, there is an increasing need for universal approaches for noninvasive prenatal screening for monogenic diseases. Here, we present a cost-effective, generic cell-free fetal DNA (cffDNA) haplotyping approach to scan the fetal genome for the presence of inherited monogenic diseases.

Methods: Families participating in the preimplantation genetic testing for monogenic disorders (PGT-M) program were recruited for this study. Two hundred fifty thousand single-nucleotide polymorphisms (SNPs) captured from maternal plasma DNA along with genomic DNA from family members were massively parallel sequenced. Parental genotypes were phased via an available genotype from a close relative, and the fetal genome-wide haplotype and copy number were determined using cffDNA haplotyping analysis based on estimation and segmentation of fetal allele presence in the maternal plasma.

Results: In all families tested, mutational profiles from cffDNA haplotyping are consistent with embryo biopsy profiles. Genome-wide fetal haplotypes are on average 97% concordant with the newborn haplotypes and embryo haplotypes.

Conclusion: We demonstrate that genome-wide targeted capture and sequencing of polymorphic SNPs from maternal plasma cell-free DNA (cfDNA) allows haplotyping and copy-number profiling of the fetal genome during pregnancy. The method enables the accurate reconstruction of the fetal haplotypes and can be easily implemented in clinical practice.

Genetics in Medicine (2020) 22:962–973; <https://doi.org/10.1038/s41436-019-0748-y>

Keywords: noninvasive prenatal testing; NIPD; cfDNA; genome-wide haplotyping; NIPS

INTRODUCTION

The discovery of cell-free fetal DNA (cffDNA) in the maternal plasma¹ has spurred the development of noninvasive prenatal screening for aneuploidies (NIPS-A). The advent of massively parallel sequencing technologies enabled noninvasive screening for the most common fetal aneuploidies (trisomy 21, 18, and 13) with high accuracy, leveraging rapid worldwide implementation in routine prenatal care.^{2,3} NIPS-A became popular because it can be applied from 10 weeks of pregnancy, reducing the risk of procedure-related miscarriage and technical challenges associated with invasive prenatal testing. Its accuracy largely outperforms the traditional first and second trimester risk assessment tests, and its implementation has resulted in a significant drop in invasive procedures.^{4,5}

In addition to aneuploidy detection, monogenic diseases can be identified by cffDNA analysis.^{6–8} Although the incidence of single-gene disorders is estimated to be about 1% of all live births⁹ and over 7000 monogenic diseases are known,¹⁰ noninvasive prenatal screening for monogenic

disorders (NIPS-M) has only been performed on a limited number of pregnancies for a small panel of genes. Different methods have been developed,^{11–15} but currently none have been widely adopted in clinical practice. While the detection of paternally inherited alleles is straightforward,¹⁶ analysis of the maternally inherited allele has been hampered by the excess of maternal DNA in cell-free DNA (cfDNA). Analytical approaches that allow noninvasive diagnosis of maternally inherited dominant or autosomal recessive monogenic diseases focus on the determination of which allele the fetus has inherited. The relative mutation dosage (RMD) by digital polymerase chain reaction (PCR) measures the relative proportions of the mutant and wild-type alleles in the maternal plasma.¹¹ However, a major disadvantage is that allele-specific probes are required and, as a consequence, are only suitable for the detection of a single targeted variant per test.^{17,18} Haplotype-based methods, including relative haplotype dosage (RHDO), deduce the fetal genotype by measuring the relative counts of alleles on haplotype blocks linked

¹Center for Human Genetics, KU Leuven, Leuven, Belgium; ²Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of Sao Paulo, Sao Paulo, Brazil; ³Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, The Netherlands; ⁴Department of Genetics and Cell Biology, GROW School for Oncology and Developmental Biology, Maastricht University, Maastricht, The Netherlands. Correspondence: Joris Robert Vermeesch (joris.vermeesch@kuleuven.be)

Submitted 1 September 2019; accepted: 27 December 2019

Published online: 6 February 2020

with the mutant allele and wild-type allele in the maternal plasma.¹⁹ Lo et al.¹³ and Kitzman et al.¹⁴ demonstrated that genome sequencing of maternal plasma DNA to 65-fold and 78-fold coverage allows the deduction of a genome-wide genetic and mutational profile of the fetus, opening opportunities to detect virtually all inherited monogenic diseases using a single platform. However, the high cost and the intensive computational analyses required for genome examinations currently prevent wide-scale clinical implementation. Target-based haplotyping methods are limited to tailored genes; a series of probes for the targeted capture of single-nucleotide polymorphisms (SNPs) flanking a particular locus needs to be selected a priori and requires 200- to 1000-fold depth of cfDNA sequencing.^{6,20,21} Hence, this approach requires disease-specific workup and cannot be universally applied for the generic diagnosis of monogenic disorders.

We previously developed a genome-wide single-cell haplotyping method, coined haplarithmisis, which enables concurrent haplotype and copy-number determination.²² Haplarithmisis is a generic method that uses informative loci from parental haplotypes across all chromosomes and assigns embryo B-allele frequencies to localize meiotic recombination sites and to measure the copy of inherited parental haplotype. This method has been clinically implemented for comprehensive embryo preimplantation genetic testing (PGT) for both monogenic disorders and aneuploidy.²³ Here, we tailor the approach for noninvasive prenatal haplotyping, and validate the method on families who underwent preimplantation genetic testing for monogenic disorders (PGT-M) where embryo haplotypes and newborn haplotypes of the uterine-transferred embryos are available. We demonstrate the feasibility of cfDNA-based haplotyping as a generic method for noninvasive prenatal detection of inherited monogenic diseases and aneuploidy detection.

MATERIALS AND METHODS

Study design

In total, nine families at risk for dominant or recessive disorders following PGT-M were included in this study (Table 1). The haplotype and mutational status of embryos were determined during routine PGT-M workflow at UZ Leuven hospital, and a healthy embryo was transferred. Genomic DNA (gDNA) from the family members, including mother, father, and a close relative (either an affected offspring or parents of the couple) was collected. Maternal plasma cfDNA was later obtained from the pregnant woman following PGT-M and embryo transfer (Supplementary Materials and Methods). In three families, approval to sample the neonate was provided and DNA was obtained. In addition, one family with a trisomy 21 child was included to create spike-in DNA samples, simulating the fetal fraction observed in maternal plasma, and evaluating the performance of the method to detect aneuploidy. The workflow of this study is illustrated in Supplementary Fig. S1.

This study was approved by the local Ethical Committee of the University Hospital Leuven (S59324). Women with a

Table 1 Clinical information of all families that followed preimplantation genetic testing (PGT).

Family	Disease	Gene	Cytogenetic location	Genomic coordinates	MI	Variant	Mother	Father	Other family member	Fetus	Phasing	Fetal sex	Gestation age (weeks)
1_181	Ceroid lipofuscinosis, neuronal 1 (OMIM 256730)	PPT1 (NM_000310.3)	1p34.2	chr1: 40,538,381–40,563,141	AR		Heterozygous c.451C>T (p.Arg151*)	Heterozygous c.566C>T (p.Pro189Leu)	Sibling combined heterozygous (c.451C>T and c.566C>T)	Noncarrier	Affected offspring	F	12
2_186	Gaucher disease, type 1 (OMIM 230800)	GBA (NM_000157.3)	1q22	chr1:155,204,238–155,214,652	AR		Heterozygous c.1448T>C (p.Leu483Pro)	Heterozygous c.1448T>C (p.Leu483Pro)	Sibling homozygous c.1448T>C (p.Leu483Pro)	Noncarrier	Affected offspring	F	12
3_085	Mitochondrial DNA depletion syndrome 6 (hepatocerebral type) (OMIM 236810)	MPV17 (NM_002437.4)	2p23.3	chr2:27,532,359–27,545,968	AR		Heterozygous c.293C>T	Heterozygous c.293C>T	Sibling homozygous c.293C>T and c.293C>T	Carrier of maternal variant			
4_158	Breast-ovarian cancer, familial, 2 (OMIM 612555)	BRCA2 (NM_000059.3)	13q13.1	chr13:32,889,616–32,973,808	AD		Noncarrier	Noncarrier	Sibling heterozygous c.116C>G (p.Thr39Arg)	Noncarrier	Affected offspring	M	12
5_149	Breast-ovarian cancer, familial, 2 (OMIM 612555)	BRCA2 (NM_000059.3)	13q13.1	chr13:32,889,616–32,973,808	AD		Noncarrier	Heterozygous c.5645C>A (p.Ser1882*)	Paternal grandmother heterozygous c.5645C>A (p.Ser1882*)	Noncarrier	Paternal parents	F	12
6_150	Neurofibromatosis type 1 (OMIM 162200)	NF1 (NM_000267.3)	17q11.2	chr17:29,421,944–29,704,694	AD		Noncarrier	Heterozygous c.7096_7101delAACTTT	Paternal grandmother heterozygous c.7096_7101delAACTTT	Noncarrier	Paternal parents	M	13
7_054	Pachyonychia congenita (OMIM 167200)	KRT16 (NM_005557.3)	17q21.2	chr17:39,766,030–39,772,151	AD		Noncarrier	Heterozygous c.374A>G (p.Asn125Ser)	Paternal grandfather heterozygous c.374A>G (p.Asn125Ser)	Noncarrier	Paternal parents	M	12
8_078	Exostoses multiple type 1 (OMIM 133700)	EXT1 (NM_000127.2)	8q24.11	chr8:118,806,729–119,124,092	AD		Heterozygous c.2133_2151del	Noncarrier	Maternal grandmother heterozygous c.2133_2151del	Noncarrier	Maternal parents	M	12
9_074	Charcot-Marie-Tooth disease type 1 (OMIM 118220)	CMT1A syndrome region including PMP22	17p12-p11	chr17:14,097,915–15,470,903	AD		Carrier of the common ~1.5-Mb duplication	Noncarrier	Maternal grandmother carrier of the common ~1.5-Mb duplication	Noncarrier	Maternal parents	M	12

AD autosomal dominant, AR autosomal recessive, F female, M male, MI mode of inheritance, XLD X-linked dominant.

successful pregnancy following PGT-M were recruited at the UZ Leuven Hospital, with informed consent.

DNA library preparation and targeted sequencing

DNA libraries were prepared using the SureSelect^{XT}HS Target Enrichment System for Illumina Paired-End Sequencing (Agilent Technologies, Santa Clara, CA). Genomic DNA was processed according to manufacturer's recommendations. In the case of cfDNA samples, between 5 and 20 ng were used for input, and the number of cycles for prehybridization PCR was optimized to 11 to generate 500–1000 ng of DNA libraries. Unique molecular identifiers (UMI) were added to DNA fragments before PCR amplification. End repair and A-tailing, ligation, and sample purification steps were performed following manufacturer's instructions. DNA libraries were hybridized to a 45-Mb custom capture library, targeting 250,000 SNPs including ~250 disease regions, subtelomeric and pericentromeric regions, and sex chromosomes. The capture library was designed based on the HumanCytoSNP-12 BeadChip (Illumina, San Diego, CA) using the Agilent SureSelect DNA Advanced Design Wizard (<https://earray.chem.agilent.com/suredesign/>) with 2×x tiling density, most stringent masking, and max performance boosting. Following hybridization and successful amplification, postcapture libraries were evaluated on Agilent 4200 TapeStation system (Agilent Technologies) using High Sensitivity D1000 SCreeTape. Concentrations were also measured by Qubit HS dsDNA Assay kit (ThermoFisher, Waltham, MA) before pooling. Pools were clustered using an Illumina cBot and sequenced with paired-end 150 reads on an Illumina NextSeq500 in high output mode. The DNA of three newborns was sequenced on an Illumina HiSeq4000 machine without UMI barcode.

Sequencing alignment and variant calling

Quality of the paired-end sequencing data was checked by FastQC v0.10.1.²⁴ Sequencing reads were aligned to GRCh37 with decoy sequences included (hs37d5) by BWA-MEM v0.7.17²⁵ and UMI barcodes were transferred to bam file as RX tag. Duplicates were marked using Picard MarkDuplicates (Broad Institute) with the awareness of UMI barcodes. Read pairs that mapped to the same genomic location and with identical molecular barcodes were grouped and ranked by base quality. Read pairs with the highest score from each molecular barcode family were kept, and PCR duplicates were removed. Low quality mapping reads (<20) and secondary alignment were filtered for downstream analyses. We used Genome Analysis Toolkit²⁶ (GATK) software suite to perform variant calling. HaplotypeCaller was used to call variants from family gDNA samples jointly and parental genotypes were phased by PhaseByTransmission. Maternal plasma samples were handled separately, and allele counts were collected using ASEReadCounter by counting paired-end fragments requiring the overlapping bases to be identical, minimal mapping quality greater than 20, and base quality greater than 2.

Only sites with more than 30 total alleles in maternal plasma were used for analysis.

cffDNA haplarithmis

The principles of cffDNA haplotyping are presented in Fig. 1. Briefly, by targeted sequencing of genomic regions genome-wide for family members and maternal plasma cfDNA, genotypes and allele counts are determined for captured SNPs. Parental genotypes are phased via an available genotype derived from a close relative, either an affected child or parents of the couple as previously described.²² The parental genotypes are divided into five groups based on paternal and maternal allele combinations (Supplementary Materials and Methods). A SNP locus is defined as informative when the genotype of one parent is heterozygous and the other is homozygous for this SNP. The informative SNPs are categorized as paternal or maternal. An informative SNP is defined as paternal when the father's genotype is heterozygous, and the mother's genotype is homozygous. Similarly, an informative SNP is defined as maternal when the mother's SNP genotype is heterozygous and the father's SNP genotype is homozygous. These paternal and maternal informative SNP loci are then subcategorized (P1, P2 or M1, M2) according to the informative phased parental SNP genotypes (Fig. 1a).

To deduce the fetal haplotype from cfDNA, we infer the alleles that originate from the fetus. For the paternal SNP category, we can easily infer the paternally inherited alleles in the fetus that differ from the maternal background alleles present in the cfDNA (Fig. 1b). For the maternal SNP category, we cannot straightforwardly distinguish the maternally inherited allele of the fetus from cfDNA as both maternal alleles are present. Nevertheless, the maternal allele inherited by the fetus will be overrepresented in maternal plasma compared with the untransmitted allele. Thus, we based the haplotyping of the fetal genome on the determination of the fetal allele ratio (FAR) that measures the proportion of fetal allele in cfDNA. First, the fetal fraction (FF) is calculated by dividing the number of reads that exhibit a paternal specific allele by the total number of reads using type 1 SNPs (Supplementary Materials and Methods). Then the FAR values are measured for SNPs where either parent has a heterozygous SNP genotype. The FAR from consecutive informative SNPs is segmented for each SNP subcategory (P1, P2 or M1, M2) separately and then jointly interpreted, defining the haplotype blocks inherited from paternal H1 and H2 or maternal H1 and H2, and pinpointing homologous recombination sites between the parental H1 and H2 (Fig. 1b). FF is used as a standard for segmented FAR value to determine homolog inheritance and to quantify copy number (Supplementary Materials and Methods).

cffDNA haplotyping validation and performance

To validate cffDNA haplotyping, we matched the cfDNA-derived haplotypes to neonatal haplotypes and array-based single-cell embryo haplotypes. Both mutational status and genome-wide haplotype concordance were compared and

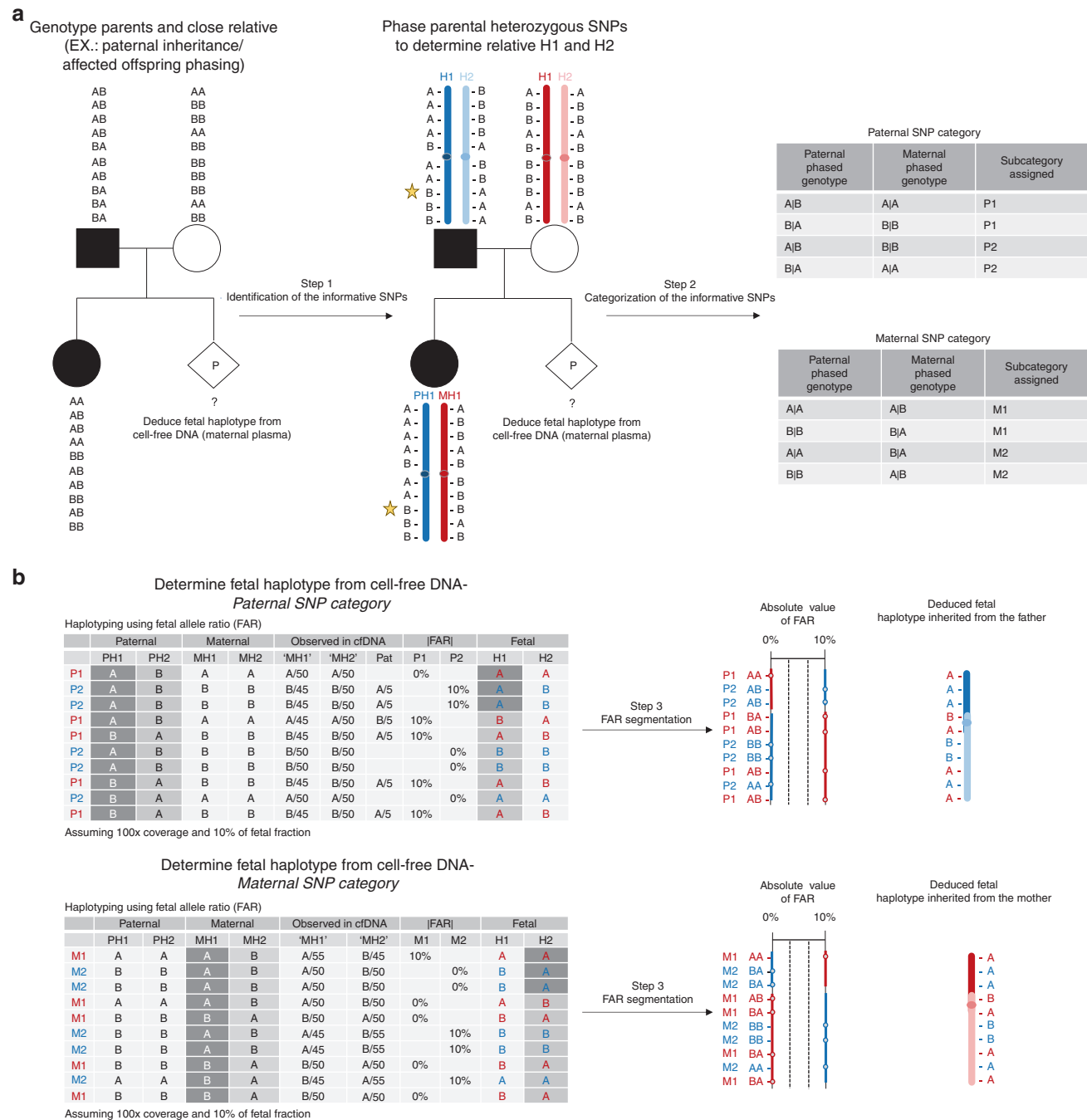


Fig. 1 Principles of cell-free fetal DNA (cffDNA) haplotyping. (a) Example of a family with an autosomal dominant disorder. DNA from the parents and the affected offspring is first genotyped and on the basis of the affected child's genotype, parental single-nucleotide polymorphisms (SNPs) can be phased to determine the transmission of paternal and maternal homologs, including the mutant allele. Paternal informative SNPs, defined as heterozygous in the father and homozygous in the mother, are identified as a step 1 phasing rule. The paternal homolog that is transmitted to the affected child must contain the causative variant and is denoted homolog 1 (H1), whereas paternal H2 carries the normal allele. Subsequently (step 2), informative SNPs are categorized to define parental SNP subcategories—P1 and P2 for paternal SNPs, and M1 and M2 for maternal SNPs. (b) Determination of fetal haplotype inheritance was based on fetal allele ratio (FAR) metric. Red and blue indicate paternal P1 and P2 SNP subcategories, and the same color code is also applied to distinguish M1 and M2 SNP subcategories. Segmentation on FAR values (step 3) was performed to define the haplotype blocks derived from paternal H1 and H2 or maternal H1 and H2, thus indicating homologous recombination sites between the parental H1 and H2.

measured. To assess the performance of the cffDNA haplotyping under effects of different factors, downsampling and simulation analyses were performed. Details are described in Supplementary Materials and Methods.

Aneuploidy detection

Synthetic spike-in samples were made to simultaneously infer fetal haplotype and detect aneuploidy. Mixed samples were created by combining 20% and 10% of DNA from the affected

proband with 80% and 90% of DNA from the mother, respectively. Chromosomal abnormalities result in FAR value deviation from the expected FF. Statistical *t* test was performed to measure shifting patterns. The haplotyping result and nonhomologous disjunction from cfDNA were confirmed with the proband DNA (Supplementary Materials and Methods).

RESULTS

Noninvasive prenatal screening for monogenic disorders

Three families had an affected offspring, and six had parents of the couple available to phase the parental genotypes (either paternal or maternal, depending who is the carrier of the variant). It is possible to infer both the paternal and maternal haplotype inheritance with an affected offspring for phasing. Phasing with parents of either father or mother determines the origin of the mutant allele in either the father or the mother (Supplementary Fig. S2) and only inheritance of the paternal or maternal haplotype is deduced. As a general observation from raw FAR values of informative SNPs (Supplementary Fig. S3), the maternal inheritance of homologous chromosome segments is more difficult to deduce visually due to the overwhelming maternal DNA background. However, we enable haplotyping of the maternally inherited genome of the fetus by FAR segmentation.

We verified the mutational profile derived from cfDNA haplotyping against the newborn profile in the first instance. In three families (families 1_181, 4_158 and 6_150), we determined neonatal haplotypes following targeted sequencing. Family 1_181 presents an autosomal recessive disorder, in which unaffected parents are heterozygous carriers for a variant in the same gene (Table 1). Haplotyping of the cfDNA identifies the paternal and maternal haplotype blocks linked with the wild-type allele at the locus of the *PPT1* gene, indicating that the fetus is not at risk for the disease. The haplotype obtained from bulk DNA of the newborn child using conventional familial analysis confirmed accurate haplotyping-based NIPS-M and concordant positioning of homologous recombination sites (Fig. 2a and Supplementary Materials and Methods). For the other two families (families 4_158 and 6_150), the disease is autosomal dominant, and the father carries the variant. Paternal parents were used for phasing. The presence of a paternal haplotype block linked with the wild-type allele confirmed the transfer of unaffected embryo and is concordant with the newborn child haplotype (Supplementary Fig. S4a, b).

For all nine families, we validated the mutational profile determined from cfDNA to the embryo biopsy haplotyping results from PGT SNP array-based haplotype analysis. In all cases the results were concordant (Fig. 2a–c and Supplementary Fig. S4a–f). Note that for family 2_186, two autosomal recessive diseases were investigated in a single PGT. Using cfDNA haplotyping, we ascertained the absence of paternal and maternal haplotype linked with the mutant alleles for type 1 Gaucher disease. The cell-free fetal haplotype revealed the same haplotype as in embryo, shown as a carrier

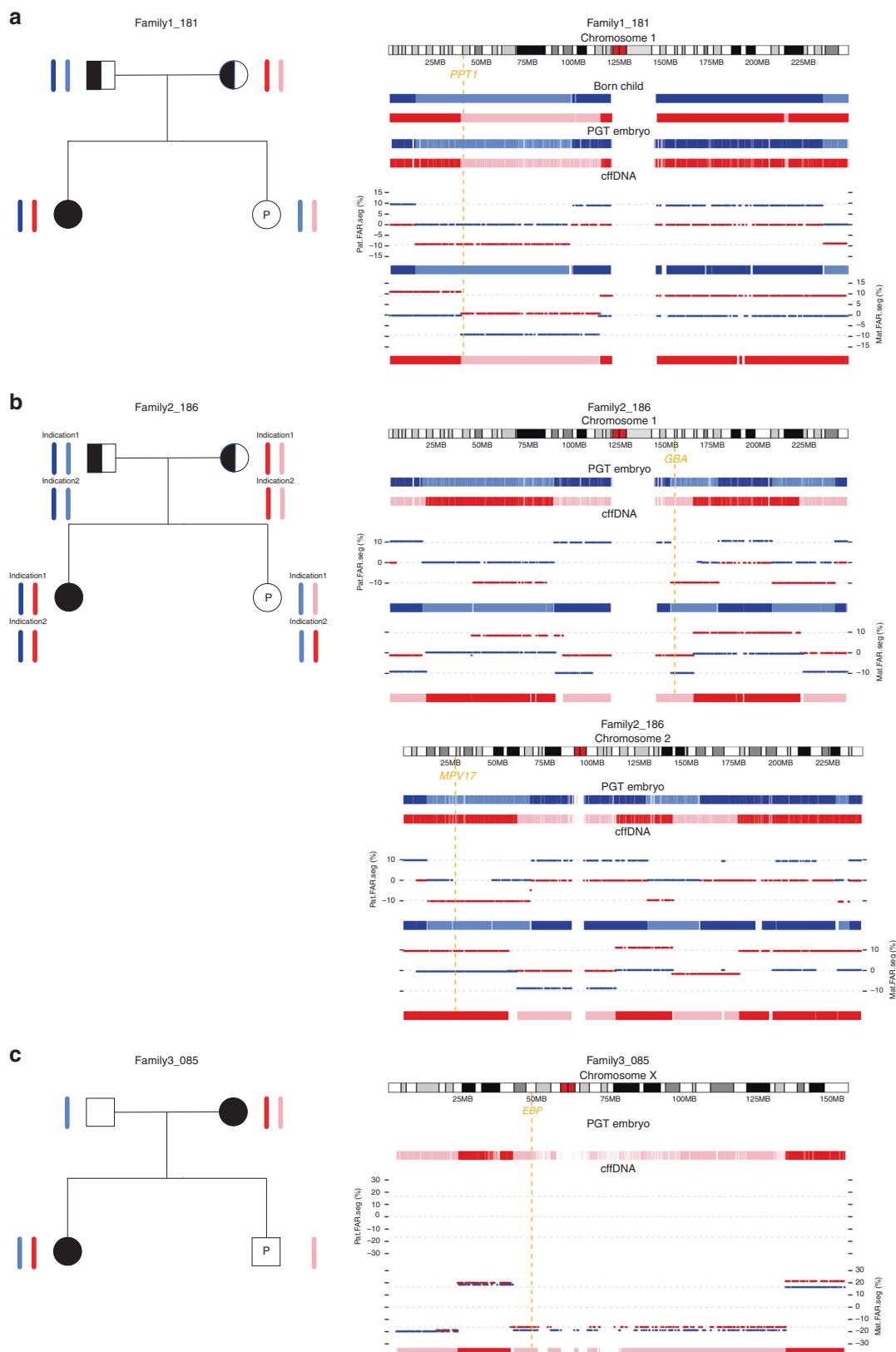
of maternal variant for mitochondrial DNA depletion syndrome 6. (Fig. 2b). Family 3_085 presents with an X-linked dominant disorder. Since the mother is the carrier of the variant in the family and the fetus is male, only the maternal haplotype inheritance is displayed in Fig. 2c.

Genome-wide cfDNA haplotyping accuracy

To evaluate the overall performance of the method, we determined the accuracy of genome-wide cfDNA haplotyping by comparing the results to conventional haplotypes derived from DNA analysis of neonatal blood when available and to single-cell haplotypes of the transferred embryo, following PGT-M. The haplotype blocks derived from born children or single cells were considered as references and haplotype blocks derived from maternal plasma DNA were matched to the reference (Supplementary Fig. S5a). Compared with neonatal genotypes, paternal and maternal informative SNPs could be deduced with 95.17% and 65.84% accuracy respectively for a 9.5% FF sample, when the prediction is only based on locus specific raw allele counts. With the use of haplotypes, the paternal and maternal genotype inference accuracy increases to 99.7% and 95.64%, respectively. Haplotyping accuracy is reduced near homologous recombination sites (Supplementary Fig. S5b,c). Comparison of cfDNA haplotypes with newborn haplotypes (Fig. 3a and Supplementary Fig. S6a) and with embryo haplotypes (Fig. 3b and Supplementary Fig. S6b) both showed an average of 99% paternal and 95% maternal haplotype concordance (Supplementary Table S1).

Above 98% accuracy was achieved for paternal haplotyping regardless of FF and sequencing depth in all cases, while maternal haplotyping accuracy varied from 90% to 97% and is FF and sequence depth dependent. Reduced haplotype resolution could be observed near the recombination sites where the maternal haplotypes' accuracy drops below 95% in a region of about 400 Kb, whereas the paternal region of lower accuracy near crossovers ranges between 100 and 350 Kb (Supplementary Table S1). Overall, haplotype accuracy and crossover resolution are impacted by FF, density of informative SNPs, and sequencing depth (Supplementary Fig. S5a). By simulating different fetal DNA fractions shown in cfDNA and maintaining the median coverage at a fixed 85-fold, the effect of FF on haplotyping accuracy and crossover was mapped. Paternal haplotypes were almost invariant to FF. In contrast, the accuracy of the maternally inherited haplotype as well as the resolution near crossovers were both greatly affected by FF. While with 20% FF, the maternal haplotypes were more than 98% accurate when compared with the reference haplotype at a homologous recombination resolution of 200 kb, and the concordance decreased to less than 80% and a resolution of 1 Mb with an FF of 5.5% (Fig. 3c).

However, it was possible to further improve the results when an affected offspring was used for phasing. In such families, SNPs that are the heterozygous in both father and mother (type 4 SNPs) were applied to improve maternally inherited haplotypes. We converted such ambiguous type 4



SNPs to unambiguous phased maternal SNPs after resolving paternal haplotypes. Adding these extra SNPs to the inference of the maternally inherited haplotype improved the accuracy when FF is low (Supplementary Fig. S7). Though we

anticipated more accurate estimation of FARs and thus improved haplotyping accuracy by raising sequencing depth especially when FF is low, we showed that even at low sequencing depth robust fetal haplotypes could be obtained.

Fig. 2 Cell-free fetal DNA (cffDNA) haplotyping analysis for monogenic disorders. Disease locus associated chromosome cffDNA haplotyping result. For each subfigure, family pedigree information is displayed together with haplotyping. The reference haplotypes from either the born child and/or from the embryo blastomere and cffDNA haplotyping results are shown. Blue in haplotype plots indicates paternal haplotype inheritance and red indicates maternal haplotype inheritance. For cffDNA haplotyping results, both segmented fetal allele ratio (FAR) values and derived haplotype blocks are shown. In segmented FAR values track, the red dotted line represents segmented P1 or M1 FAR and blue for segmented P2 or M2 FAR, and the distance between P1 and P2 or M1 and M2 segmentation in the same genomic region indicates fetal fraction. We flipped FAR values of P1 subcategory around 0 and FAR values of M1 were subjected to less than or equal to 0 in visualization for clear separation between informative single-nucleotide polymorphism (SNP) subcategories. Disease loci are indicated by a yellow vertical line. (a) cffDNA haplotype compared with neonate haplotype and embryo haplotype. The paternal homolog carrying the variant is represented in dark blue and maternal homolog carrying the variant is represented in dark red. The disease locus resided in the light color block of both paternal and maternal haplotypes, indicating wild-type alleles were transmitted. (b, c) cffDNA haplotyping results compared with the embryo haplotype. (b) Two disease indications of the family are shown. (c) Inheritance of an X-linked disorder is shown. PGT preimplantation genetic testing.

To further investigate the effect of sequencing depth on haplotyping, we downsampled two cfDNA-sequencing samples having 16.5% FF and 9.5% FF. Although the range of sequencing depth was 50-fold to 96-fold in our samples (Supplementary Table S1), downsampling simulations showed that cfDNA haplotyping performance was stably maintained at 40-fold sequencing depth for 16.5% FF and was only reduced to below 95% concordance at 30-fold sequencing depth (Fig. 3d). With FF at 9.5%, haplotyping accuracy also only dropped significantly when sequencing depth was reduced to below 30% of original depth (Supplementary Fig. S8).

Aneuploidy detection using cffDNA haplotyping

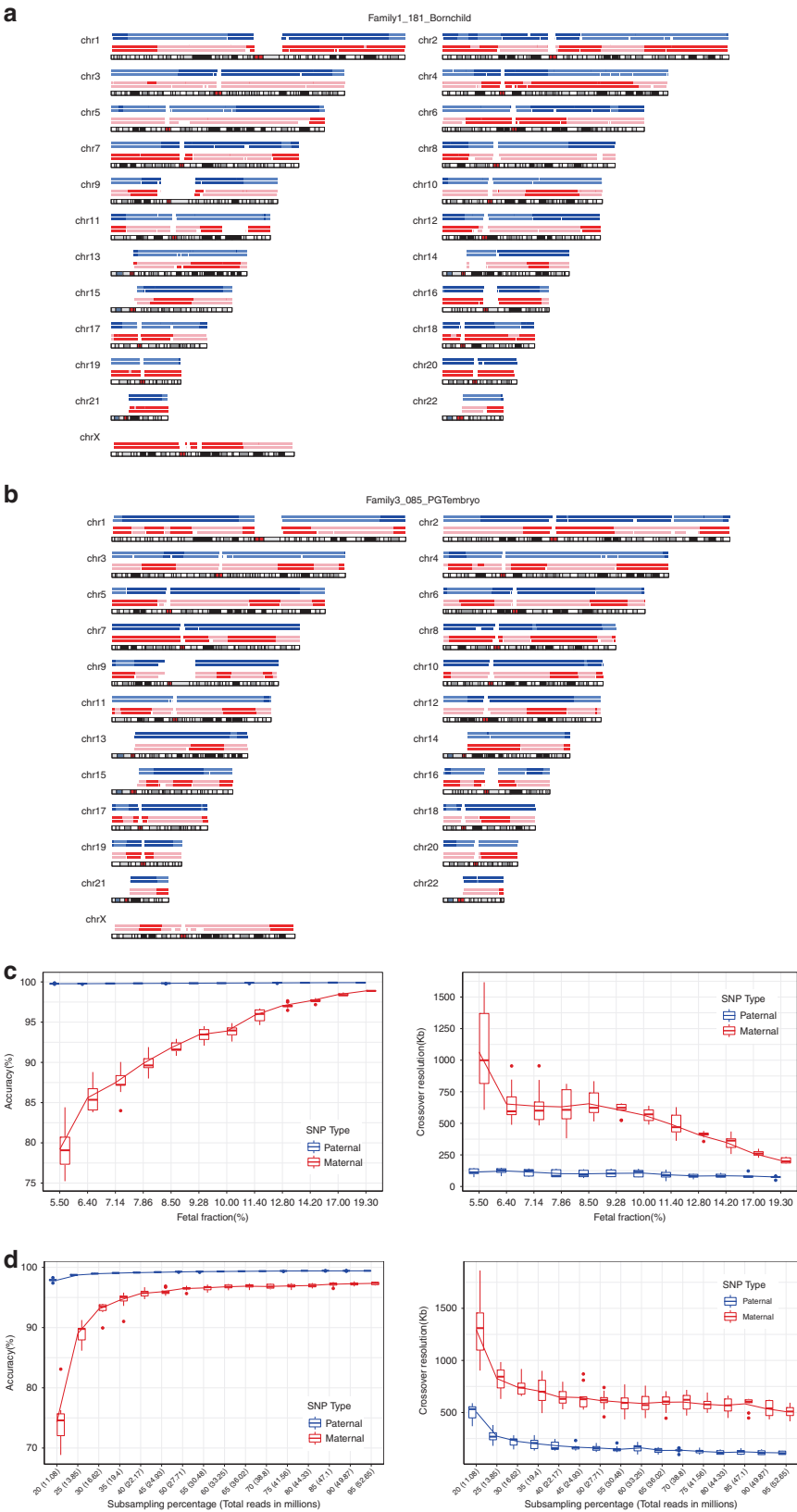
In addition to inherited monogenic disease detection (NIPS-M), we explored the capability of the methodology to detect simultaneously aneuploidy (NIPS-A) using synthetic spike-in samples. As illustrated in Supplementary Fig. S9, a trisomy would lead to a deviation of the segmented FAR value from the expected FF. In a maternally inherited trisomy, the segmented FAR value of the paternal SNPs shifts only marginally; however, maternal FAR values will shift systematically away from the FF value of the diploid autosomes. For instance, assuming 10% FF and 100-fold coverage, we would expect maternal FAR value to shift to 4.76% or 14.29% rather than 0% or 10%. While for a paternally inherited trisomy, both paternal and maternal FAR values deviate from the expected FF (Supplementary Fig. S9a, b). Accordingly, our data revealed that the paternal FAR values on chromosome 21 presented close to the expected normal FF levels, while the maternal FAR values for both M1 and M2 subcategories shifted away from the expected FF, being near to the theoretical trisomy FAR values. From the *t* test, maternal (M1 and M2) mean FAR values of chromosome 21 showed significant difference from their corresponding subcategorical mean FAR values of other chromosomes (Supplementary Fig. S10). As a result, a maternally inherited trisomy 21 was determined (Fig. 4a). The predicted haplotypes have both maternal haplotypes present, indicating that the trisomy is the consequence of a maternal meiotic nondisjunction. Sequencing data from the proband also confirmed this and showed a concordant phasing between the spike-in samples and the child's DNA (Fig. 4b). Despite trisomy 21, the genome-wide

haplotyping demonstrated 99.9% and 99% accuracy for paternal and maternal inheritance for 20% FF spike-in sample, and 99.8% and 97% for 10% FF spike-in sample.

DISCUSSION

We demonstrate that genome-wide targeted capture and sequencing of polymorphic SNPs from maternal cfDNA along with parental and additional family member's DNA allows haplotyping and copy-number profiling of the fetal genome during pregnancy. cffDNA haplotype analysis enables the accurate reconstruction of the fetal haplotypes without the need for deep sequencing or genome sequencing analyses. A wide spectrum of monogenic disorders and aneuploidy are readily detectable via this approach. This opens the venue for concurrent NIPS-M and NIPS-A. With uptake of testing and technology refinement, detection of subchromosomal aneuploidy and copy-number detection will become feasible as well.

We envision cffDNA haplotype analysis to be a universal NIPS-M that avoids the necessity to design specific panels defining particular loci to be analyzed. Cost is one of the major factors that limit the scalability of NIPS-M. The capture design and targeted sequencing used in this method can make NIPS-M more affordable in the long term. In contrast to the RHDO method, where more than 200-fold coverage of the target loci is required, the method leverages segmentation of fetal allele ratio over multiple informative SNPs, allowing a significant reduction of the required sequencing depth. Samples can be multiplexed to further reduce costs. To enhance the haplotype inference accuracy, unique molecular identifiers were incorporated to reduce amplification artifacts and technical biases were removed by using multiple filtering criteria, monitoring sequencing errors and applying dynamic bias corrections. We set standard classification rules based on FF to assure sufficient evidences supporting the homolog assignment (Supplementary Materials and Methods). A range of conditions, including dominant, recessive, and X-linked monogenic diseases can be assessed in this generic non-invasive prenatal diagnosis test. Multiple variants can be identified in one test without variant-specific designs, as shown in the case of family 2. In case of aneuploidies, the parental origin and the segregation error (meiotic or mitotic) can be deduced.



Parental haplotypes are deduced from the genotypes of other family members. Sometimes, those relatives are not available. Direct parental haplotyping through long-read²⁰ or

linked-read^{6,27} technology can offer a solution to haplotype inference of the family without additional family members. In the longer haul, the availability of population haplotypes will

Fig. 3 Genome-wide cell-free fetal DNA (cffDNA) haplotyping accuracy. (a) Genome-wide comparison of cffDNA haplotyping results with the neonate haplotype from family 1_181. For each chromosome, dark and light blue represent paternal haplotyping and dark and light red represent maternal haplotyping; the upper haplotype track refers to born child haplotype and lower track represents cell-free DNA (cfDNA)-based fetal haplotype. (b) Genome-wide comparison of cffDNA haplotypes to the embryo blastomere haplotype from family 3_085. The upper track shows the single-cell haplotype and lower track represents cfDNA-derived haplotype. (c) Simulation of the impact of fetal fraction (FF) on haplotyping accuracy and resolution. (d) Effect of sequencing depth on the performance of cffDNA haplotyping for 16.5% FF.

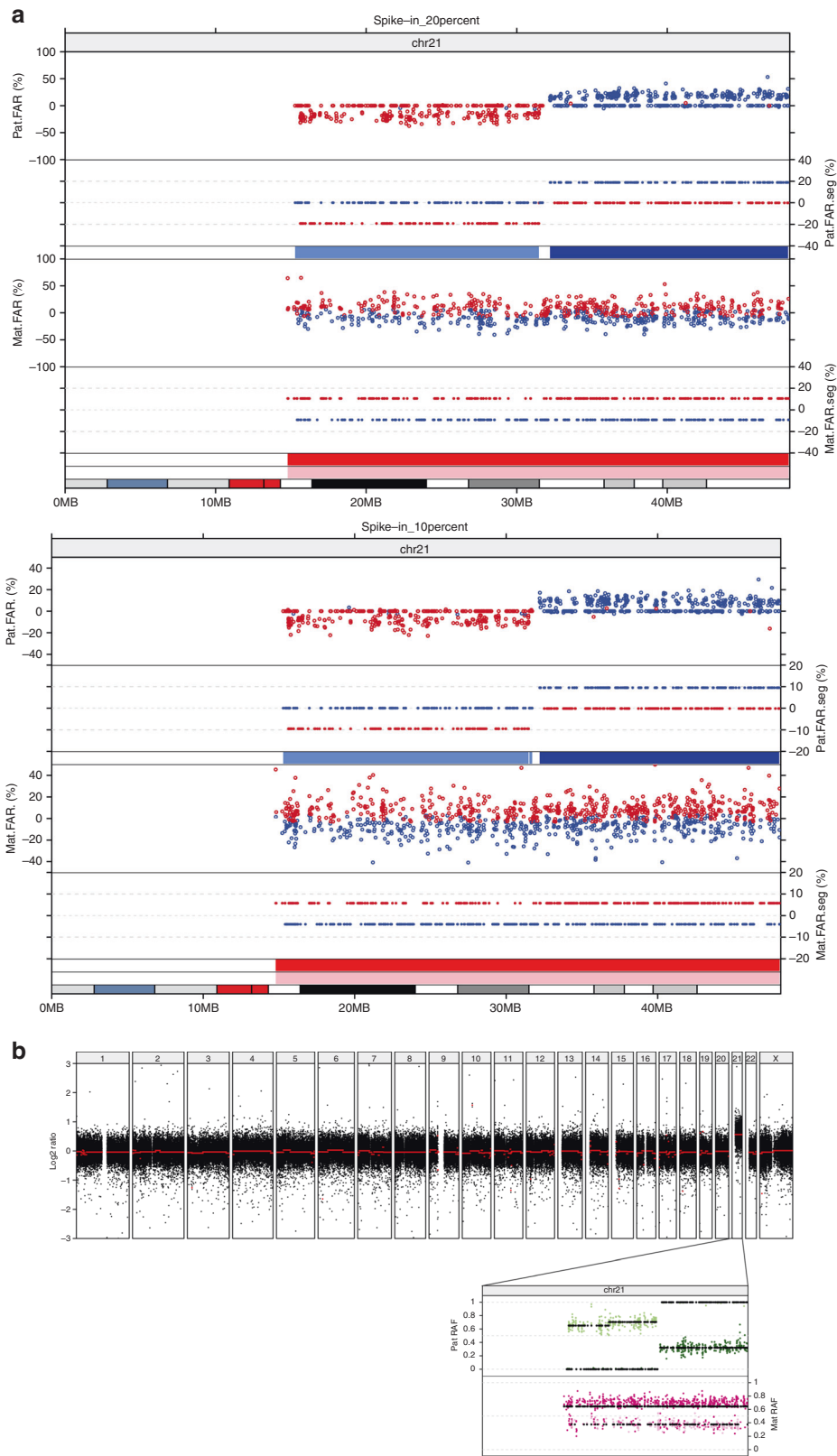
allow inference of the disease allele, especially for the most common recessive disorders.^{28,29} Those haplotypes could be imputed to reduce the need for parental and grandparental haplotyping. Also, although the method was designed to be generically applicable to monogenic diseases detection, it is not suitable for de novo variant detection. Zhang *et al.*³⁰ demonstrated a capture design of the most frequent dominant disorders for the detection of de novo and paternally inherited disease-causing variants. It might become possible to add capture probes to the current design. However, the approach should be compensating for the high sequencing depth required for de novo variant detection compared with the relative lower sequencing depth required here. We reached an overall 97% concordance with embryo and neonatal haplotypes and the discordance arises mainly from the homologous recombination regions. As a general limitation, meiotic homologous recombinations occurring near the mutant gene would not allow inference of whether the fetus is a carrier of variant or not. In such case, an invasive test can be recommended. The accuracy of the homologous recombination and haplotype construction are determined by the interplay of the fetal fraction, the density of informative SNPs, and the sequencing depth around the genomic region. Low fetal fraction leads to reduced accuracy particular for the inference of the maternally inherited haplotype, but this may be remedied by higher density of informative SNPs. Though we yielded conclusive results for all clinical cases presented in the study, from simulation we estimated that to get an overall accuracy in maternal inheritance haplotyping above 90% requires 7.8% fetal fraction with moderate sequencing depth. Paternally inherited haplotypes can readily be detected even when the fetal fraction is 3% at about 85-fold coverage (Supplementary Fig. S11). Of note, we have one case where the mother is a carrier of ~1.5-Mb duplication causing Charcot-Marie-Tooth disease type 1 syndrome. Though maternal copy-number variations may interfere with FAR estimations, the method showed tolerance for the excessive maternal allele background for relatively small duplication. As we can already detect the maternal copy number in this case (Supplementary Fig. S12a, b), parental copy numbers of larger sizes are very likely to be detected, and the region can be taken into consideration for proper interpretation of results. Placental chromosomal mosaicism might be another factor affecting the analysis. From PGT-A of blastocysts it is becoming clear that many blastocysts carry aneuploidies in a fraction of the cells. Although we demonstrated that the impact of embryonic aneuploidy seems to be marginal during

prenatal development,³¹ placental mosaic aneuploidies have been reported in NIPS-A.^{32,33} It remains to be determined whether mosaic aneuploidies would interfere with this approach.

Generic methods to haplotype and profile aneuploidies in embryos have transformed preimplantation genetic testing for monogenic diseases (PGT-M) and are becoming an integral aspect of in vitro fertilization procedures.^{34,35} Because of the risk for a spontaneous pregnancy during the PGT procedure and possible laboratory procedure errors,³⁶ a prenatal diagnostic test (chorionic villus sampling or amniocentesis) is currently highly recommended to confirm the transfer of an unaffected embryo. Since fetal genetic testing for monogenic disorders currently requires an invasive procedure that may have miscarriage risk,³⁷ most families refrain from undergoing the test. cffDNA haplarithmis represents a safer alternative for these families. As proof-of-concept, we actually performed targeted sequencing on families who underwent PGT-M by genome-wide haplotyping and demonstrated a very high concordance of the embryo single-cell and cfDNA derived haplotypes. In families who undergo PGT-M by haplotyping, NIPS-M can be streamlined as one workflow where phasing of parental genotypes has already been performed in the PGT-M process. Hence, noninvasive prenatal fetal haplotyping would require only analysis of the targeted cfDNA.

Genetic carrier screening has been offered to individuals and couples based on family history or ethnic background. Screening for cystic fibrosis and thalassemia is recommended and has been rolled out for general preconception and prenatal populations.³⁸ Moreover, advances in next-generation sequencing and better understanding of disease-causing variants continuously drive expansion of screening panels.³⁹ With an increasing number of genetic disorders recognized to be practical for screening, growing awareness that each individual can be a carrier of variants that may cause recessive disorders, and increasing use of carrier screening in the general population,⁴⁰ new approaches to reduce the transmission of disease alleles that lead to severe morbidity and mortality are desirable. Therefore, our method could be applied in combination with carrier screening programs to help couples who are at high risk for inherited diseases but who cannot use, do not want to use, or do not have access to preimplantation genetic testing make autonomous reproductive decisions.

In summary, haplarithmis makes noninvasive genome-wide fetal haplotyping and aneuploidy detection with targeted



sequencing accessible to all. This universal cffDNA haplotyping approach could easily be adopted by genetic testing laboratories and would provide comfort to both the couples

and the caretakers involved. Following this proof-of-concept study, we expect expanded clinical studies to further validate the method more precisely.

Fig. 4 Aneuploidy detection using spike-in samples. (a) Maternally inherited trisomy 21 detected in spike-in DNA samples with 20% and 10% fetal fraction, respectively. In the 20% spike-in sample, maternal fetal allele ratio (FAR) values dropped around 10% and -10%, indicating an extra copy from the “fetus” on chromosome 21. Similarly, in the 10% FF, maternal FAR values were around 5% and -5%. (b) Validation of maternally inherited trisomy case using sequencing data from the proband. The copy-number plot (log2 ratio) indicated chromosomal copy number on chromosome 21, with black dot as log2 ratio of each target and red line as segmented value. The paternal and maternal reference allele ratio from chromosome 21 also deviates from standard reference allele ratio of 0.5 and the homologous recombinations are consistent with the result derived from spike-in samples.

SUPPLEMENTARY INFORMATION

The online version of this article (<https://doi.org/10.1038/s41436-019-0748-y>) contains supplementary material, which is available to authorized users.

DATA AVAILABILITY

All sequencing data from nine families who underwent PGT and one family with a trisomy 21 proband have been submitted to the European Genome-phenome Archive (<https://www.ebi.ac.uk/ega/>) under accession number EGAS00001003634.

CODE AVAILABILITY

Code is available on request.

ACKNOWLEDGEMENTS

We are grateful to all families who participated in this study. This work was funded by Agilent, by KULeuven funding (C1/018), the Horizon 2020 WIDENLIFE: 692065 to J.R.V. and T.V., and by the Fonds voor Wetenschappelijk Onderzoek (FWO-Vlaanderen: G080217N) to J.R.V. D.V. was supported by FAPESP (2017/23448-8).

DISCLOSURE

The authors declare no conflicts of interest. M.Z.E., T.V. and J.R.V. co-invented haplarithmisis (‘Haplotyping and copy-number typing using polymorphic variant allelic frequencies’) licensed to Agilent Technologies.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

- Lo YMD, et al. Presence of fetal DNA in maternal plasma and serum. *Lancet*. 1997;350:485–487.
- Bayindir B, et al. Noninvasive prenatal testing using a novel analysis pipeline to screen for all autosomal fetal aneuploidies improves pregnancy management. *Eur J Hum Genet*. 2015;23:1286–1293.
- Bianchi DW, Chiu RWK. Sequencing of circulating cell-free DNA during pregnancy. *N Engl J Med*. 2018;379:464–473.
- Song K, Musci TJ, Caughey AB. Clinical utility and cost of noninvasive prenatal testing with cfDNA analysis in high-risk women based on a US population. *J Matern Fetal Neonatal Med*. 2013;26:1180–1185.
- Hui L, Hutchinson B, Poulton A, Halliday J. Population-based impact of noninvasive prenatal screening on screening and diagnostic testing for fetal aneuploidy. *Genet Med*. 2017;19:1338–1345.
- Hui WWI, et al. Universal haplotype-based noninvasive prenatal testing for single gene diseases. *Clin Chem*. 2017;63:513–524.
- New MI, et al. Noninvasive prenatal diagnosis of congenital adrenal hyperplasia using cell-free fetal DNA in maternal plasma. *J Clin Endocrinol Metab*. 2014;99:E1022–E1030.
- Xu Y, et al. Haplotype-based approach for noninvasive prenatal tests of Duchenne muscular dystrophy using cell-free fetal DNA in maternal plasma. *Genet Med*. 2015;17:889–896.
- Fujikura K. Global carrier rates of rare inherited disorders using population exome sequences. *PLoS One*. 2016;11:e0155552.
- Bell CJ, et al. Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci Transl Med*. 2011;3:65ra4–65ra4.
- Lun FMF, et al. Noninvasive prenatal diagnosis of monogenic diseases by digital size selection and relative mutation dosage on DNA in maternal plasma. *Proc Natl Acad Sci U S A*. 2008;105:19920–19925.
- Chiu RWK, Cantor CR, Lo YMD. Noninvasive prenatal diagnosis by single molecule counting technologies. *Trends Genet*. 2009;25:324–331.
- Lo YMD, et al. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci Transl Med*. 2010;2:61ra91–61ra91.
- Kitzman JO, et al. Noninvasive whole-genome sequencing of a human fetus. *Sci Transl Med*. 2012;4:137ra76–137ra76.
- Fan HC, et al. Noninvasive prenatal measurement of the fetal genome. *Nature*. 2012;487:320–324.
- Verhoeve TI, et al. Noninvasive prenatal diagnosis (NIPD) for single gene disorders: cost analysis of NIPD and invasive testing pathways. *Prenat Diagn*. 2016;36:636–642.
- Tsui NBY, et al. Noninvasive prenatal diagnosis of hemophilia by microfluidics digital PCR analysis of maternal plasma DNA. *Blood*. 2011;117:3684–3691.
- Barrett AN, McDonnell TCR, Chan KCA, Chitty LS. Digital PCR analysis of maternal plasma for noninvasive detection of sickle cell anemia. *Clin Chem*. 2012;58:1026–1032.
- Chiu EKL, Hui WWI, Chiu RWK. cfDNA screening and diagnosis of monogenic disorders—where are we heading? *Prenat Diagn*. 2018;38:52–58.
- Vermeulen C, et al. Sensitive monogenic noninvasive prenatal diagnosis by targeted haplotyping. *Am J Hum Genet*. 2017;101:326–339.
- Parks M, et al. Noninvasive prenatal diagnosis of spinal muscular atrophy by relative haplotype dosage. *Eur J Hum Genet*. 2017;25:416–422.
- Zamani Esteki M, et al. Concurrent whole-genome haplotyping and copy-number profiling of single cells. *Am J Hum Genet*. 2015;96:894–912.
- Dimitriadou E, et al. Principles guiding embryo selection following genome-wide haplotyping of preimplantation embryos. *Hum Reprod*. 2017;32:687–697.
- Andrews S. FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. Accessed 27 Jan 2020.
- Li H, and Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26:589–595.
- McKenna A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–1303.
- Zheng GXY, et al. Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nat Biotechnol*. 2016;34:303–311.
- Browning SR, Browning BL. Haplotype phasing: existing methods and new developments. *Nat Rev Genet*. 2011;12:703–714.
- Zeevi DA, et al. Proof-of-principle rapid noninvasive prenatal diagnosis of autosomal recessive founder mutations. *J Clin Invest*. 2015;125:3757–3765.
- Zhang J, et al. Noninvasive prenatal sequencing for multiple Mendelian monogenic disorders using circulating cell-free fetal DNA. *Nat Med*. 2019;25:439–447.
- Zamani Esteki M, et al. In vitro fertilization does not increase the incidence of de novo copy number alterations in fetal and placental lineages. *Nat Med*. 2019;25:1699–1705.
- Brisson N, et al. Predicting fetoplacental chromosomal mosaicism during noninvasive prenatal testing. *Prenat Diagn*. 2018;38:258–266.

33. van der Meij KRM, et al. TRIDENT-2: national implementation of genome-wide non-invasive prenatal testing as a first-tier screening test in the Netherlands. *Am J Hum Genet.* 2019.
34. Vermeesch JR, Voet T, Devriendt K. Prenatal and preimplantation genetic diagnosis. *Nat Rev Genet.* 2016;17:643–656.
35. Handyside AH. ‘Designer babies’ almost thirty years on. *Reproduction.* 2018;156:F75–F79.
36. Wilton L, Thornhill A, Traeger-Synodinos J, Sermon KD, Harper JC. The causes of misdiagnosis and adverse outcomes in PGD. *Hum Reprod.* 2009;24:1221–1228.
37. Salomon LJ, Sotiriadis A, Wulff CB, Odibo A, Akolekar R. Risk of miscarriage following amniocentesis or chorionic villus sampling: systematic review of literature and updated meta-analysis. *Ultrasound Obstet Gynecol.* 2019;54:442–451.
38. Bajaj K, Gross SJ. Carrier screening: past, present, and future. *J Clin Med.* 2014;3:1033–1042.
39. Lizarin GA, Haque IS. Expanded carrier screening: a review of early implementation and literature. *Semin Perinatol.* 2016;40:29–34.
40. Haque IS, et al. Modeled fetal risk of genetic diseases identified by expanded carrier screening. *JAMA.* 2016;316:734–742.