Genetics in Medicine

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

--Manuscript Draft--

Manuscript Number:	GIM-D-19-00723R1
Article Type:	Article
Section/Category:	Reproductive Genetics
Keywords:	Non-invasive prenatal diagnosis; cell free DNA; haplotyping; monogenic diseases
Corresponding Author:	Joris R. Vermeesch, Ph.D. Centre for Human Genetics, KU Leuven Leuven, BELGIUM
First Author:	Huiwen Che, M.S.
Order of Authors:	Huiwen Che, M.S.
	Darine Villela
	Eftychia Dimitriadou
	Cindy Melotte
	Nathalie Brison
	Maria Neofytou
	Kris Van Den Bogaert
	Olga Tsuiko
	Koen Devriendt
	Eric Legius
	Masoud Zamani Esteki
	Thierry Voet
	Joris Robert Vermeesch
Manuscript Region of Origin:	BELGIUM
Abstract:	Purpose Whereas non-invasive prenatal testing for aneuploidies (NIPT-A) is widely implemented, there is an increasing need for universal approaches for noninvasive prenatal testing for monogenic diseases (NIPT-M). 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. 250000 SNPs captured from maternal plasma DNA along with genomic DNAs 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 (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.

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

To Dr Robert D Steiner
Editor-in-Chief
Genetics in Medicine

Ref: GIM-D-19-00723

Dear Dr. Steiner,

Thank you for your email regarding revision of the manuscript "Non-invasive prenatal diagnosis by genome-wide haplotyping of cell-free plasma DNA". We have carefully read the comments and revised the manuscript accordingly.

We would like to thank the reviewer for thoughtful reading the manuscript and for their constructive suggestions. Our responses are in blue given in a point-by-point manner (the reviewers' comments are in italics).

Sincerely,
Joris R. Vermeesch, Ph.D.
Professor of Laboratory of Cytogenetics and Genome Research
Head of the Department of Human Genetics, KU Leuven
Herestraat 49, box 602, 3000 Leuven, Belgium
e-mail: joris.vermeesch@uzleuven.be

Response to the reviewers

Reviewer #1: The potential application of NIPD to follow up the results of PGT is exciting, as increasing number of PGT couples are reluctant to follow the current recommendations for invasive prenatal diagnosis; thus this manuscript presents a clinical scenario and experimental details of such a practical application. The work is well written, and will be of interest to a wide audience.

- *Perhaps the authors could expand in a more detail the accuracy of a calculation of fetal fraction, and a lower fetal fraction limit for the conclusion.*

  We added discussion of lower fetal fraction limit on page 18 line 369-374.

- *It will be also useful if authors could discuss any additional factors that may impact the accuracy of NIPD. One of such factors that creates a concern is a genetic instability in preimplantation development that may affect the accuracy of testing. However, these are minor suggestions to consider for discussion.*

  Regarding genetic instability of preimplantation embryos, we have recently demonstrated that the impact of embryonic aneuploidy seems to diminish during prenatal development (Zamani-Esteki *et al*., 2019 *Nature Med*), further minimizing the concern that embryo genetic instability can be a confounding factor for cfDNA haplarithmisis. We commented it on page 19 line 377-385.

Reviewer #2: Huiwen Che and colleagues describe a method they devised for genome-wide haplotyping on cell-free DNA, to allow concurrent diagnosis of aneuploidies with inheritance of known familial pathogenic variants in recessive, dominant and X-linked disorders. They use as underlying principle for their approach a method which they named "haplarithmisis" and that was original developed for PGT on in vitro-fertilized preimplantation embryos.

This is primarily a proof of concept paper with initial technical validation, but the sample size is fairly small (9 families and one "artificially mixed" aneuploidy sample. Overall the paper is well written and clear, supporting tables and figures (including in supplemental data) are appropriate and informative, with some details that need to be addressed (see below).

It is not the first demonstration that this type of data can be obtained from cffDNA, and the approach labeled "haplarithmisis" and determination of Fetal Allele Ratio (FAR) is essentially based on RHDO of multiple SNPs at once, with some refinements to optimize analysis by optimizing which SNPs are used in the analysis (addition of informative type 4 SNPs) and pairing with circular binary segmentation that allows them to overcome noise and an have estimate of recombination sites in the

haplotypes.

The conceptual innovation in this paper results from the combined aneuploidy and monogenic disease testing possibilities on a more genome-wide basis using a less costly and faster method than those previously published that also avoids the requirement for disease-specific work-up.

Although the numbers are small, the validation for the haplotyping on pregnancies conceived after PGT-M, wherein both embryo and newborn haplotypes are available for comparison is elegant. I am less impressed with the fairly limited approach to the validation of the aneuploidy detection, which is entirely based on one mixing experiment with spike-in samples, since they always have the limitations that fragmentation is not identical to that of naturally occurring cfDNA. The study would have more value if a few cases where analysis is done on cffDNA from aneuploid pregnancies with comparison of data from newborn genomic DNA/karyotype (or from products of conception results). The ability to determine parental origin of the aneuploidy if validated is an advantage of this method.

We thank the Reviewer for an extensive evaluation and agree that mixing experiment with spike-ins is a limitation, when it comes to validation of aneuploidy detection. Unfortunately, we were not able to perform direct comparison of aneuploidy detection in cffDNA and affected newborn or product of conception, as in the framework of this project we did not have an aneuploidy clinical case. However, we are keeping running the clinical study to collect samples to further validate the method. We foresee that we can have access to aneuploidy cases to overcome the limitations of 'artificial' sample.

I have the following additional detailed comments and suggestions:

- *The abstract is clear and reflects the data and conclusions of the manuscript accurately. I like the introduction of NIPT-A and NIPT-M, as terms that nicely parallel PGT-A and PGT-M. Maybe authors could consider including NIPT-SR (or NIPT-CN or "copy number"?) in their terminology as well and elaborate on this suggested terminology in the discussion some more?*

  We address the point in the discussion on page 16 line 319-322.

- *Introduction: on P 3, first paragraph: NIPT-A indeed resulted in a significant drop on invasive procedures, however the word "unnecessary", which is an interpretation, should be removed from this sentence. Concern has been raised that while procedures for aneuploidy testing have gone down, the side effect is that there is also a decline in procedures for diagnoses that are not found by NIPT-A and that would have been detected by a diagnostic procedure.*

We removed “unnecessary” from the sentence on page 3 line 53.

- *Introduction: P4, last line (and also in methods): can authors define the word "haplarithmisis". I realize the method has been previously published and is briefly described, but it will help readers to understand what the basis for this unusual terminology is.*

  We added explanations for haplarithmisis on page 5 line 83-85.

- *I suggest that for all relevant figures you consider different color combinations to designate the different parental SNP subcategories and haplotype blocks besides red and blue. Sometimes red blue differentiates two different paternal haplotypes of the same parent and sometimes it differentiates maternal versus paternal in the fetal DNA.*

  We clarified the color pattern in the figure legend.

- *Figure 3: Figure 3a is a little hard to follow. Consider more clarification in legend or revision of this panel. Figures 3c and 3d are too small. I had to significantly enlarge these on the computer monitor to actually see the two blue and two red tracks for each chromosome. It is impossible to see this on the printed version. This should be corrected (same applies for similar figures in supplemental data).*

  We moved Figure 3a and 3b to Supplementary Figure S5a and S5b and added supplementary figure legend for S5a. We rearranged Figure 3 and Supplementary Figure S6 to present figures properly.

- *I would like to see more clearly described if the actual variant (wild-type versus carrier) was detected accurately in addition to correct haplotype establishment. The authors must have some data on this.*

  The variants of the clinical samples are not in the target design and thus unfortunately we do not have the information of the allele counts on the specific variant locus. We expect that if actual variants happen to be within the capture design, we can obtain locus-specific information to support haplotyping results.

- *There should be some comment in the discussion how the 97% concordance with embryo and newborn haplotypes would affect clinical utility if introduced for clinical application.*

  We commented the clinical utility of the method in the discussion on page 18 line 361-365.

- *In the discussion, the authors state that their method is better than working up individual cases and doing targeted cfDNA analysis for familial variants. That statement should be supported by a discussion that considers comparison of cost, time estimates to results, accuracy, and the limitations of their new method. In particular, bespoke testing has been shown to be diagnostic (not screening) in certain studies while their assay seems less accurate for maternally inherited variants and has problems if there is meiotic recombination near a pathogenic variant.*

  We expanded detailed discussion regarding the advantage and limitations of the method on page 16 line 324-344 and page 17 line 346-385. The maternally inherited variants show lower accuracy compared with paternally inherited variants due to the maternal background DNA in plasma samples. The haplotyping of maternal inheritance is also largely affected by fetal fraction. If the variant lies in meiotic recombination region, the method cannot reach a conclusive result and invasive test is required.

- *Considering they also suggest this as a potential non-invasive follow-up assay for PGT-M confirmation, an accurate test to confirm would be desired. I would be concerned that for an individual case, a specific error or limitation in a PGT-M assay done by haplarithmisis may well be carried through if haplarithmisis is also used as the basis for the NIPT-M test. Can they comment on this? maybe bespoke non-invasive testing is the way to go.*

  Computational workflow behind haplarithmisis for PGT-M has been extensively validated during the developmental stage and has been successfully implemented into the routine clinical practice since 2015 (Dimitriadou *et al.*, 2017, *Hum Reprod*). Although haplotyping-based NIPT-M is based on haplarithmisis, the computation of NIPT-M relies on fetal proportion of cfDNA is independent of single-cell B-allele frequency in PGT-M. The major limiting factor that might prohibit the use of this technology both in PGT-M and NIPT-M is the proximity of the locus of interest to homologous recombination site.

- *The other limitation, the inability to detect de novo variants is touched on very briefly but is significant. Have the authors explored their data for the ability to detect de novo variants on the paternal allele? There should be some data as they are in essence similar to type 1 variants.*

  Although the test allows detection of paternal *de novo* mutations, the design did not focus on *de novo* mutations. As suggested by the reviewer, important disease-causing genes can be included into the capture design in the future and screening for *de novo* mutations can be incorporated with inherited variants identification. This is commented on page 17 line 354-365.

- *In the discussion, the comment on the 1% risk for procedure related pregnancy loss for an invasive procedure is an overestimate. Recent data suggest that it is in fact much lower than 1%, with several papers indicating there is no significant difference between amniocentesis and CVS. This statement must be revised. Please note that there is a more recent meta-analysis by the same authors of reference 37, which itself is incorrectly referenced as it does not say that the risk is 1%.*

  We removed the incorrect number. The text was rephrased on page 19 line 394, and the reference was updated to Salomon *et al*., 2019 *Ultrasound Obstet Gynecol*.

- *The authors have a case with maternal CMT1A and comment that it does not interfere with the analysis, but what the resolution of this method would be to detect such CN changes in cfDNA is not addressed. Can they comment on this? Can they model it?*

  We added maternal copy number profile of this case to the Supplementary Figure 12 and we comment further on this case in the discussion on page 19 line 377-380.

- *The comments on carrier screening should qualify that the most desirable approach would be preconception carrier screening followed by PGT-M rather than carrier testing during pregnancy followed by diagnostic testing or NIPT-M*

  This item was addressed on page 20 line 413-414.

- *It would be nice if the discussion could end with a comment about wider clinical validation to determine clinical utility.*

  We revised the discussion to address the item on page 21 line 420-422.

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

Huiwen Che[1], Darine Villela[1,2], Eftychia Dimitriadou[1], Cindy Melotte[1], Nathalie Brison[1], Maria Neofytou[1], Kris Van Den Bogaert[1], Olga Tsuiko[1], Koen Devriendt[1], Eric Legius[1], Masoud Zamani Esteki[1,3,4], Thierry Voet[1], Joris Robert Vermeesch[1*]

[1]Center for Human Genetics, KU Leuven, Leuven, Belgium

[2]Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of Sao Paulo, Sao Paulo, Brazil

[3]Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, The Netherlands

[4]Department of Genetics and Cell Biology, GROW School for Oncology and Developmental Biology, Maastricht University, Maastricht, The Netherlands

*Corresponding author: Joris Robert Vermeesch, Center for Human Genetics, University Hospital Leuven, KU Leuven, Herestraat 49, box 602, 3000 Leuven, Belgium.
Telephone: +32 16 34 5941
Email: joris.vermeesch@kuleuven.be

ABSTRACT

Purpose

Whereas non-invasive prenatal testing for aneuploidies (NIPT-A) is widely implemented, there is an increasing need for universal approaches for noninvasive prenatal testing for monogenic diseases (NIPT-M). 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. 250000 SNPs captured from maternal plasma DNA along with genomic DNAs 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 (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.

INTRODUCTION

The discovery of cffDNA in the maternal plasma[1] has spurred the development of non-invasive prenatal genetic testing (NIPT). The advent of massively parallel sequencing technologies enabled non-invasive screening for the most common fetal aneuploidies (trisomy 21, 18 and 13) with high accuracy, which leverage rapid worldwide implementation of NIPT-A in routine prenatal care[2,3]. NIPT-A became popular because it can be applied from 10 weeks of pregnancy, reduces the risk of procedure-related miscarriage and technical challenges associated with invasive prenatal testing. The accuracy largely outperforms the traditional first and second trimester risk assessment tests. Its implementation resulted in a significant drop of ~~unnecessary~~ 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 life births[9] and over 7000 monogenic diseases are known[10], non-invasive prenatal testing for monogenic disorders 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 has 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 cfDNA. Analytical approaches that allow non-invasive 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 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 one single targeted variant per test[17,18]. Haplotype-based methods deduce the fetal genotype by measuring the relative counts of alleles on haplotype blocks linked with the mutant allele and wild-type allele in the maternal plasma[19]. Lo *et al.*[13] and Kitzman *et al*.[14] demonstrated that whole-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 one single platform. However, the high cost and the intensive computational analyses required for whole 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 SNPs flanking a particular locus needs to be selected *a priori* and require 200- to 1000-fold depth of cfDNA sequencing[6,20,21]. Hence, this approach requires disease specific work-up 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 use informative loci from parental haplotypes across all chromosomes and assign 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 pre-implantation genetic testing (PGT) for both monogenic disorders and aneuploidy[23]. Here, we tailor the approach for non-invasive prenatal haplotyping, and validated the method on families that underwent PGT-M where embryo haplotypes and newborn haplotypes of the uterine transferred embryos are available. We demonstrate the feasibility of cffDNA based haplotyping as a generic method for non-invasive 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 DNAs from the family members, including mother, father, and a close relative (either an affected offspring or parents of the couple) were 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 evaluate the performance of the method to detect aneuploidy. The workflow of this study is illustrated in **Supplementary Figure S1**.

This study was approved by the local Ethical Committees of the University Hospital Leuven (S59324). Women with a 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^{XTHS} Target Enrichment System for Illumina Paired-End Sequencing (Agilent Technologies, California, USA). Genomic DNAs were processed according to manufacturer's recommendations. In the case of cfDNA samples, between 5-20 ng were used for input, and the number of cycles for pre-hybridization 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, which targets 250,000 SNPs that includes ~250 disease regions, subtelomeric and pericentromeric regions, and sex chromosomes. The capture library was designed based on the HumanCytoSNP-12 BeadChip (Illumina, California, USA) using the Agilent SureSelect DNA Advanced Design Wizard (https://earray.chem.agilent.com/suredesign/) with 2x tiling density, most stringent masking, and max performance boosting. Following hybridization and successful

amplification, post-capture 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, Massachusetts, USA) before pooling. Pools were clustered using an Illumina cBot and sequenced with paired-end 150 reads on an Illumina NextSeq500 in high output mode. Three newborns DNA were sequenced on an Illumina HiSeq4000 machine without UMI barcode.

Sequencing alignment and variant calling

Quality of the pair-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 v0.7.17[25] and UMI barcodes were transferred to bam file as RX tag. Duplicates were marked using Picard MarkDuplicates (the 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 families were kept, and PCR duplicates were removed. Low quality mapping reads (< 20) and secondary alignment were filtered for downstream analyses. We used Genome Analysis Toolkit[26] (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 pair-end fragments requiring the overlapping bases to be identical, minimal mapping quality greater than 20, and base quality

146 greater than 2. Only sites with more than 30 total alleles count in maternal plasma were used
147 for analysis.
148
149 **cffDNA haplarithmisis**
150 The principles of cffDNA haplotyping are presented in **Figure 1**. Briefly, by targeted
151 sequencing genomic regions genome-wide for family members and maternal plasma cfDNA,
152 genotypes and allele counts are determined for captured SNPs. Parental genotypes are phased
153 via an available genotype derived from a close relative, either an affected child or parents of
154 the couple as previously described[22]. The parental genotypes are divided into 5 groups based
155 on paternal and maternal allele combinations (**Supplementary Materials and Methods**). A
156 SNP locus is defined informative when the genotype of one parent is heterozygous and the
157 other is homozygous for this SNP. The informative SNPs are categorized as paternal or
158 maternal. An informative SNP is defined ‘‘paternal’’ when the father’s genotype is
159 heterozygous, and the mother’s genotype is homozygous. Similarly, an informative SNP is
160 defined ‘‘maternal’’ when the mother’s SNP genotype is heterozygous and the father’s SNP
161 genotype is homozygous. These paternal and maternal informative SNP loci are then
162 subcategorized (P1, P2 or M1, M2) according to the informative phased parental SNP
163 genotypes (**Figure 1a**).
164
165 To deduce the fetal haplotype from cfDNA, we infer the alleles that originate from the fetus.
166 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 (**Figrue 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 comparing to 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 (**Figure 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 derive haplotypes to neonatal haplotypes and array-based single cell embryo haplotypes. Both mutational status and genome-wide haplotype concordance were compared and 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 the respective pairs 80% and 90% of DNA from the mother. Chromosomal abnormalities result in FAR value deviation from the expected FF. Statistical t-test was performed to measure shifting patterns. The haplotyping result and non-homologous disjunction from cfDNA were confirmed with the proband DNA (**Supplementary Materials and Methods**).

RESULTS

Non-invasive prenatal testing for monogenic disorders

Three families have an affected offspring, and six have parents of the couple available to phase the parental genotypes (either paternal or maternal, depending who the carrier of variant is). 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 Figure S2**) and only inheritance of the paternal or maternal haplotype is deduced. As a general observation from raw FAR values of informative SNPs (**Supplementary Figure 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 cffDNA haplotyping against the newborn profile in the first instance. In three families (Family 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 cffDNA 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 NIPT-M and concordant positioning of homologous recombination sites (**Figure 2a** and **Supplementary Materials and Methods**). For the other two families (Family 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 Figure S4a** and **b**).

For all nine families, we validated the mutational profile determined from cfDNA to the embryo biopsy haplotyping results from PGT SNP array-based haplarithmisis analyses. In all

cases the results were concordant (**Figure 2a-c** and **Supplementary Figure S4a-f**). Note that for Family 2_186, two autosomal recessive diseases were investigated in a single PGT. Using cffDNA 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 from embryo as being carrier of maternal variant for mitochondrial DNA depletion syndrome 6 indication (**Figure 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 **Figure 2c**.

Genome-wide cffDNA haplotyping accuracy

To evaluate the overall performance of the method, we determined the accuracy of genome-wide cffDNA 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 Figure S5a**). Comparing to 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 Figure S5b-c**).

The comparison of cffDNA haplotypes to newborn haplotypes (**Figure 3a** and **Supplementary Figure S6a**), and to embryo haplotypes (**Figure 3b** and **Supplementary Figure 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 400Kb, whereas the paternal region of lower accuracy near crossovers ranges between 100 and 350Kb (**Supplementary Table S1**). Overall, haplotype accuracy and crossover resolution are impacted by FF, density of informative SNPs and sequencing depth (**Supplementary Figure 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 200kb, the concordance decreased to less than 80% and a resolution of 1Mb with an FF of 5.5% (**Figure 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 SNPs 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. By adding these extra SNPs to the inference of the maternally inherited haplotype improved the accuracy when FF is low (**Supplementary Figure 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. 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 (**Figure 3d**). With FF at 9.5%, haplotyping accuracy also only dropped greatly when sequencing depth reduced to below 30% of original depth (**Supplementary Figure S8**).

Aneuploidy detection using cffDNA haplotyping

In addition to inherited monogenic disease detection (NIPT-M), we explored the capability of the methodology to detect simultaneously aneuploidy (NIPT-A) using synthetic spike-in samples. As illustrated in **Supplementary Figure 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 that 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 Figure S9a** and **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. Using *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 Figure S10**). As a result, a maternally inherited trisomy 21 was determined (**Figure 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 DNA (**Figure 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, additional family member's DNA, allows haplotyping and copy number profiling of the fetal genome during pregnancy. cffDNA haplarithmisis analysis enables the accurate reconstruction of the fetal haplotypes without the need for deep sequencing nor whole genome sequencing analyses. A wide spectrum of monogenic disorders and aneuploidy are readily detectable via this approach. This opens the venue for concurrent NIPT-M and NIPT-A. With uptake of the testing and technology refinement, detection of sub-chromosomal aneuploidy and copy number detection will become feasible as well.

We envision cffDNA haplarithmisis to be a universal NIPT that avoids the necessity to design specific panels defining particular loci to be analyzed. Cost is one of the major factors that limits the scalability of NIPT-M. The capture design and targeted sequencing used in this method can make NIPT-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 that allows a significant reduction of the required sequencing depth. Samples can be multiplexed to further reduce costs. ~~In contrast to current whole genome cffDNA haplotyping methods that require whole genome sequencing to achieve genome-wide fetal haplotyping resolution, targeted sequencing reduces the amount of sequencing needed and reduces the complexity of processing deep whole genome data.~~ ~~Targeted sequencing and haplotyping of a genome-wide~~

~~SNP panel in cfDNA thus lowers the overall cost and reduces processing time.~~ 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-read[20] 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 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. ~~To overcome this, amplicon sequencing of specific loci has been explored~~. 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* mutation detection as compared to the relative lower sequencing depth required here. ~~A more general limitation of the method is that meiotic homologous recombinations occurring near the mutant gene would not allow to infer whether the fetus is carrier of variant or not.~~ 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 maternally 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 Figure 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 the 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 FigureS12a-b**), parental copy number of larger sizes are very likely 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 NIPT[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 which ~~is estimated to have a 1% risk for procedure induced miscarriage,~~ may have miscarriage risk[37], most families renounce from undergoing the test. cffDNA haplarithmisis represents a safer alternative for these families. As proof-of-concept, we actually performed targeted sequencing on families that underwent PGT-M by genome-wide haplotyping and do demonstrate a very high concordance of the

398 embryo single-cell and cfDNA derived haplotypes. In families that undergo PGT-M by
399 haplotyping, NIPT-M can be streamlined as one workflow where phasing of parental
400 genotypes has already been performed in the PGT-M process. Hence, non-invasive prenatal
401 fetal haplotyping would require only analysis of the targeted cfDNA.
402
403 Genetic carrier screening has been offered to individuals and couples based on family history
404 or ethnic background. Screening of cystic fibrosis and thalassemia have been recommended
405 and is rolled out for general preconception and prenatal populations[38]. Moreover, technology
406 advances in next generation sequencing and better understanding of disease causing variants
407 continuously drive expansion of screening panels[39]. As an increasing number of genetic
408 disorders recognized to be practical for screening, the raising awareness that each individual
409 can be a carrier of variants that may cause recessive disorders and the increasing uptake of
410 carrier screening in the general population[40], new approaches to reduce the transmission of
411 disease alleles that leads to severe morbidity and mortality are desirable. In such scenario, our
412 method could be applied in combination with carrier screening program to help couples ~~at~~
413 ~~high risk for inherited diseases~~ who are at high risk for inherited diseases but cannot, or do
414 not want to or do not have access to preimplantation genetic testing to make autonomous
415 reproductive decisions.
416
417 In summary, haplarithmisis makes non-invasive genome-wide fetal haplotyping and
418 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 care-takers involved. Following this proof-of-concept study, we expect expanded clinical studies to further validate the method more precisely.

Data availability

All sequencing data from 9 families underwent PGT and 1 family with trisomy 21 proband have been submitted to the European Genome-phenome Archive (https://www.ebi.ac.uk/ega/) under accession numbers EGAS00001003634.

Code availability

Code is available in GitHub at https://github.com/vermeeschlab/cfDNAhaplotyping

Acknowledgements

We are grateful to all families that 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.V. D.V. was supported by FAPESP (2017/23448-8).

Disclosure

The authors declare no conflicts of interest.

References

1. Lo, Y. M. D. *et al.* Presence of fetal DNA in maternal plasma and serum. *The Lancet* **350**, 485–487 (1997).
2. 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.* **23**, 1286–1293 (2015).
3. Bianchi, D. W. & Chiu, R. W. K. Sequencing of Circulating Cell-free DNA during Pregnancy. *N. Engl. J. Med.* **379**, 464–473 (2018).
4. Song, K., Musci, T. J. & Caughey, A. B. Clinical utility and cost of non-invasive prenatal testing with cfDNA analysis in high-risk women based on a US population. *J. Matern. Fetal Neonatal Med.* **26**, 1180–1185 (2013).
5. 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.* **19**, 1338–1345 (2017).
6. Hui, W. W. I. *et al.* Universal Haplotype-Based Noninvasive Prenatal Testing for Single Gene Diseases. *Clin. Chem.* **63**, 513–524 (2017).
7. New, M. I. *et al.* Noninvasive Prenatal Diagnosis of Congenital Adrenal Hyperplasia Using Cell-Free Fetal DNA in Maternal Plasma. *J. Clin. Endocrinol. Metab.* **99**, E1022–E1030 (2014).
8. 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.* **17**, 889–896 (2015).

461 9. Fujikura, K. Global Carrier Rates of Rare Inherited Disorders Using Population Exome
462 Sequences. *PLOS ONE* **11**, e0155552 (2016).
463 10. Bell, C. J. *et al.* Carrier Testing for Severe Childhood Recessive Diseases by Next-
464 Generation Sequencing. *Sci. Transl. Med.* **3**, 65ra4-65ra4 (2011).
465 11. Lun, F. M. F. *et al.* Noninvasive prenatal diagnosis of monogenic diseases by digital size
466 selection and relative mutation dosage on DNA in maternal plasma. *Proc. Natl. Acad. Sci.*
467 **105**, 19920–19925 (2008).
468 12. Chiu, R. W. K., Cantor, C. R. & Lo, Y. M. D. Non-invasive prenatal diagnosis by single
469 molecule counting technologies. *Trends Genet.* **25**, 324–331 (2009).
470 13. Lo, Y. M. D. *et al.* Maternal Plasma DNA Sequencing Reveals the Genome-Wide Genetic
471 and Mutational Profile of the Fetus. *Sci. Transl. Med.* **2**, 61ra91-61ra91 (2010).
472 14. Kitzman, J. O. *et al.* Noninvasive Whole-Genome Sequencing of a Human Fetus. *Sci.*
473 *Transl. Med.* **4**, 137ra76-137ra76 (2012).
474 15. Fan, H. C. *et al.* Non-invasive prenatal measurement of the fetal genome. *Nature* **487**, 320–
475 324 (2012).
476 16. Verhoef, T. I. *et al.* Non-invasive prenatal diagnosis (NIPD) for single gene disorders: cost
477 analysis of NIPD and invasive testing pathways. *Prenat. Diagn.* **36**, 636–642 (2016).
478 17. Tsui, N. B. Y. *et al.* Noninvasive prenatal diagnosis of hemophilia by microfluidics digital
479 PCR analysis of maternal plasma DNA. *Blood* **117**, 3684–3691 (2011).
480 18. Barrett, A. N., McDonnell, T. C. R., Chan, K. C. A. & Chitty, L. S. Digital PCR Analysis
481 of Maternal Plasma for Noninvasive Detection of Sickle Cell Anemia. *Clin. Chem.* **58**,

1026–1032 (2012).

19. Chiu, E. K. L., Hui, W. W. I. & Chiu, R. W. K. cfDNA screening and diagnosis of monogenic disorders – where are we heading? *Prenat. Diagn.* **38**, 52–58 (2018).

20. Vermeulen, C. *et al.* Sensitive Monogenic Noninvasive Prenatal Diagnosis by Targeted Haplotyping. *Am. J. Hum. Genet.* **101**, 326–339 (2017).

21. Parks, M. *et al.* Non-invasive prenatal diagnosis of spinal muscular atrophy by relative haplotype dosage. *Eur. J. Hum. Genet.* **25**, 416–422 (2017).

22. Zamani Esteki, M. *et al.* Concurrent Whole-Genome Haplotyping and Copy-Number Profiling of Single Cells. *Am. J. Hum. Genet.* **96**, 894–912 (2015).

23. Dimitriadou, E. *et al.* Principles guiding embryo selection following genome-wide haplotyping of preimplantation embryos. *Hum. Reprod.* **32**, 687–697 (2017).

24. Andrews, S. *FastQC: a quality control tool for high throughput sequence data.*

25. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv13033997 Q-Bio* (2013).

26. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

27. Zheng, G. X. Y. *et al.* Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nat. Biotechnol.* **34**, 303–311 (2016).

28. Browning, S. R. & Browning, B. L. Haplotype phasing: existing methods and new developments. *Nat. Rev. Genet.* **12**, 703–714 (2011).

29. Zeevi, D. A. *et al.* Proof-of-principle rapid noninvasive prenatal diagnosis of autosomal

503 recessive founder mutations. *J. Clin. Invest.* **125**, 3757–3765 (2015).
504 30. Zhang, J. *et al.* Non-invasive prenatal sequencing for multiple Mendelian monogenic
505 disorders using circulating cell-free fetal DNA. *Nat. Med.* 1 (2019) doi:10.1038/s41591-
506 018-0334-x.
507 31. Zamani Esteki, M. *et al.* In vitro fertilization does not increase the incidence of de novo
508 copy number alterations in fetal and placental lineages. *Nat. Med.* **25**, 1699–1705 (2019).
509 32. Brison, N. *et al.* Predicting fetoplacental chromosomal mosaicism during non-invasive
510 prenatal testing. *Prenat. Diagn.* **38**, 258–266 (2018).
511 33. van der Meij, K. R. M. *et al.* TRIDENT-2: National Implementation of Genome-Wide
512 Non-Invasive Prenatal Testing as a First-Tier Screening Test in the Netherlands. *Am. J. Hum.*
513 *Genet.* (2019) doi:10.1016/j.ajhg.2019.10.005.
514 34. Vermeesch, J. R., Voet, T. & Devriendt, K. Prenatal and pre-implantation genetic diagnosis.
515 *Nat. Rev. Genet.* **17**, 643–656 (2016).
516 35. Handyside, A. H. 'Designer babies' almost thirty years on. *Reproduction* **156**, F75–F79
517 (2018).
518 36. Wilton, L., Thornhill, A., Traeger-Synodinos, J., Sermon, K. D. & Harper, J. C. The causes
519 of misdiagnosis and adverse outcomes in PGD. *Hum. Reprod.* **24**, 1221–1228 (2009).
520 37. Salomon, L. J., Sotiriadis, A., Wulff, C. B., Odibo, A. & Akolekar, R. Risk of miscarriage
521 following amniocentesis or chorionic villus sampling: systematic review of literature and
522 updated meta-analysis. *Ultrasound Obstet. Gynecol.* **54**, 442–451 (2019).
523 38. Bajaj, K. & Gross, S. J. Carrier Screening: Past, Present, and Future. *J. Clin. Med.* **3**, 1033–

524 1042 (2014).

525 39. Lazarin, G. A. & Haque, I. S. Expanded carrier screening: A review of early

526 implementation and literature. *Semin. Perinatol.* **40**, 29–34 (2016).

527 40. Haque, I. S. *et al.* Modeled Fetal Risk of Genetic Diseases Identified by Expanded Carrier

528 Screening. *JAMA* **316**, 734–742 (2016).

529

FIGURE LEGENDS

Figure 1. Principles of 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 SNPs can be phased to determine the transmission of paternal and maternal homologues, 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 homologue that is transmitted to the affected child must contain the causative variant and is denoted homologue 1 (H1), whereas paternal H2 carries the normal allele. Subsequently (step 2), informative SNPs are categorized to define parental SNPs 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 SNPs subcategories, and the same color code is also applied to distinguish M1 and M2 SNPs 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.

Figure 2. cffDNA haplotyping analysis for monogenic disorders. Disease locus associated chromosome cffDNA haplotyping result. For each sub figure, family pedigree information is displayed in together with haplotyping. The reference haplotypes from either the born child and/or from the embryo blastomere and cffDNA haplotyping results are shown. Color blue in

haplotype plots indicates paternal haplotype inheritance and color red indicates maternal haplotype inheritance. For cffDNA haplotyping results, both segmented 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 SNPs subcategories. Disease loci are indicated with a yellow vertical line. **a**, cffDNA haplotype compared to neonate haplotype and embryo haplotype. The paternal homologue carrying the variant is represented in dark blue and maternal homologue carrying the variant is represented in dark red. The disease locus resided in light color block of both paternal and maternal haplotypes, indicating wild-type alleles were transmitted. **b-c**, cffDNA haplotyping results compared to the embryo haplotype. **b**, Two disease indications of the family are shown. **c**, Inheritance of an X-linked disorder is shown.

Figure 3. Genome-wide cffDNA haplotyping accuracy. ~~**a**, Illustration of density of SNPs, sequencing depth and FF effects on haplotype accuracy and crossover resolution metrics. **b**, For a 9.5%FF cfDNA samples, homologous recombination accuracy measured in 10Kb bins that flank out from both sides of true (neonatal) crossovers, is shown. On x-axis, 0 denotes the true crossover breakpoint sites. Positive and negative distances correspond to extension to the right and to the left side of breakpoints.~~ **a**, Genome-wide comparison of cffDNA

haplotyping results to the neonate haplotype from Family 1_181. For each chromosome, dark and light blue represents paternal haplotyping and dark and light red represents maternal haplotyping; the upper haplotype track refers to born child haplotype and lower track represents cfDNA-based fetal halplotype. **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.

Figure 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 20% spike-in sample, maternal 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 spike-in samples derived result.

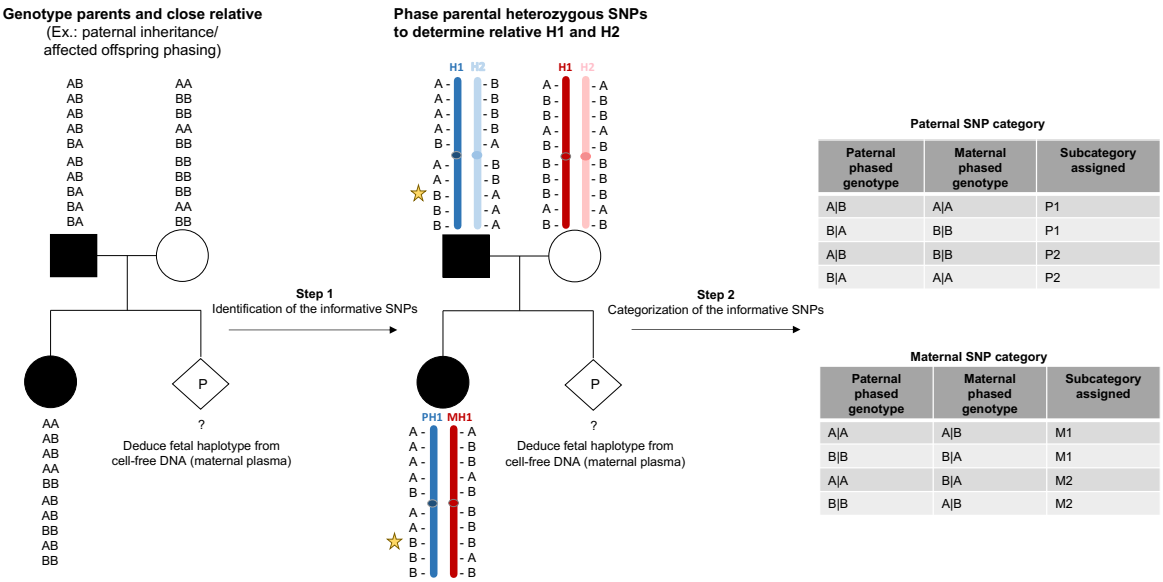
Conflict of Interest

The authors have nothing to disclose.

Figure

Figure 1

a



Paternal SNP category

Paternal phased genotype	Maternal phased genotype	Subcategory assigned
A|B	A|A	P1
B|A	B|B	P1
A|B	B|B	P2
B|A	A|A	P2

Maternal SNP category

Paternal phased genotype	Maternal phased genotype	Subcategory assigned
A|A	A|B	M1
B|B	B|A	M1
A|A	B|A	M2
B|B	A|B	M2

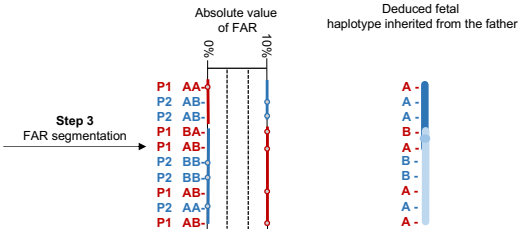
b

Determine fetal haplotype from cell-free DNA
- Paternal SNP category

Haplotyping using fetal allele ratio (FAR)

	Paternal		Maternal		Observed in cfDNA			|FAR|		Fetal	
	PH1	PH2	MH1	MH2	'MH1'	'MH2'	Pat	P1	P2	H1	H2
P1	A	B	A	A	A/50	A/50		0%		A	A
P2	A	B	B	B	B/45	B/50	A/5		10%	A	B
P2	A	B	B	B	B/45	B/50	A/5		10%	A	B
P1	A	B	A	A	A/45	A/50	B/5	10%		B	A
P1	B	A	B	B	B/45	B/50	A/5	10%		A	B
P2	A	B	B	B	B/50	B/50			0%	B	B
P2	A	B	B	B	B/50	B/50			0%	B	B
P1	B	A	B	B	B/45	B/50	A/5	10%		A	B
P2	B	A	A	A	A/50	A/50			0%	A	A
P1	B	A	B	B	B/45	B/50	A/5	10%		A	B

Assuming 100x coverage and 10% of fetal fraction



Determine fetal haplotype from cell-free DNA -
Maternal SNP category

Haplotyping using fetal allele ratio (FAR)

	Paternal		Maternal		Observed in cfDNA		|FAR|		Fetal	
	PH1	PH2	MH1	MH2	'MH1'	'MH2'	M1	M2	H1	H2
M1	A	A	A	B	A/55	B/45	10%		A	A
M2	B	B	A	B	A/50	B/50		0%	B	A
M2	B	B	A	B	A/50	B/50		0%	B	A
M1	A	A	A	B	A/50	B/50	0%		A	B
M1	B	B	B	A	B/50	A/50	0%		B	A
M2	B	B	A	B	A/45	B/55		10%	B	B
M2	B	B	A	B	A/45	B/55		10%	B	B
M1	B	B	B	A	B/50	A/50	0%		B	A
M2	A	A	B	A	B/45	A/55		10%	A	A
M1	B	B	B	A	B/50	A/50	0%		B	A

Assuming 100x coverage and 10% of fetal fraction

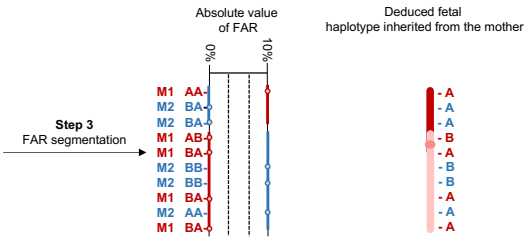
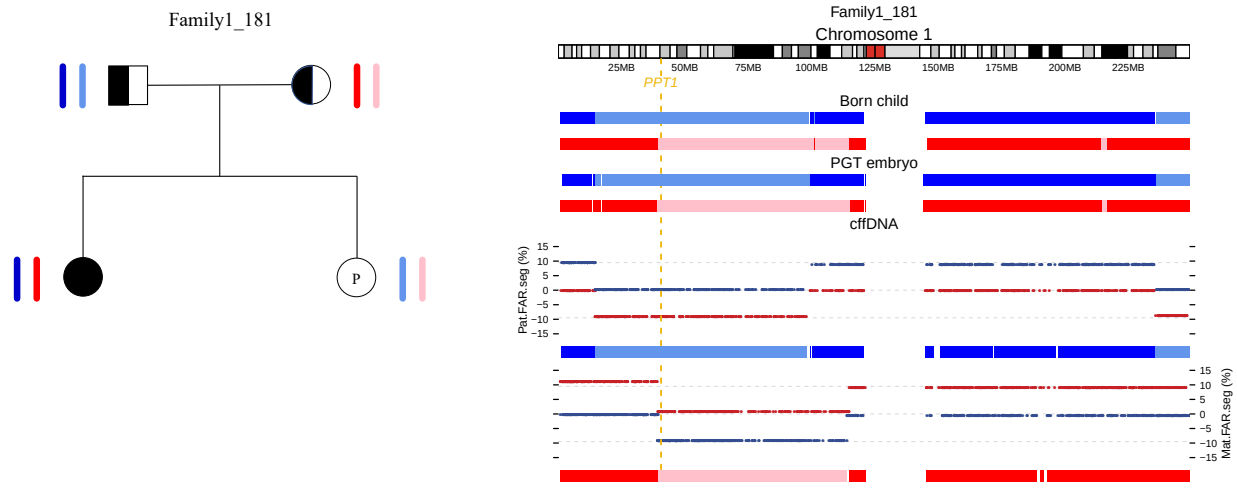
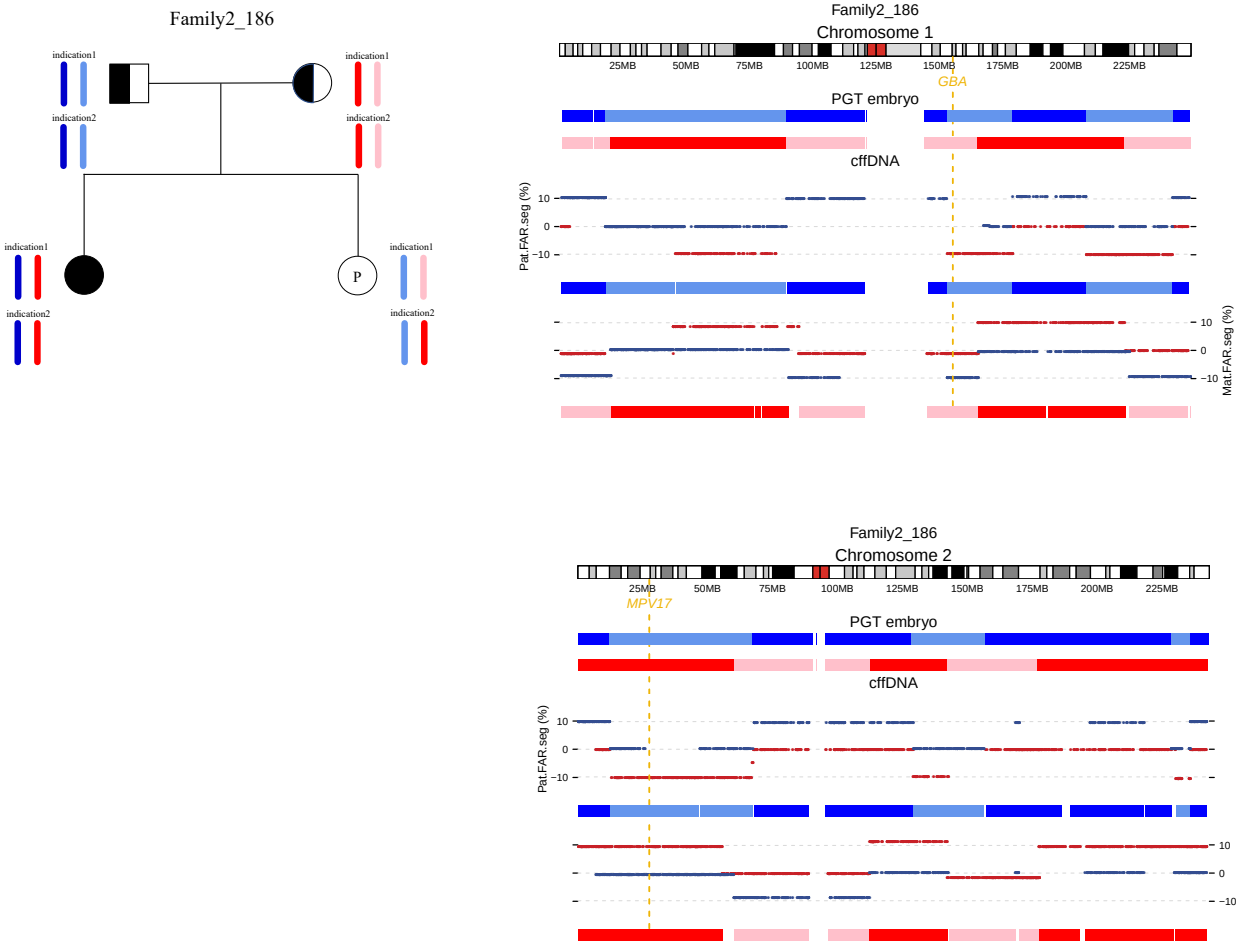


Figure 2

a


Family1_181
P
Family1_181
Chromosome 1
25MB
50MB
75MB
100MB
125MB
150MB
175MB
200MB
225MB
PPT1
Born child
PGT embryo
cffDNA
Pat.FAR.seg (%)
Mat.FAR.seg (%)

b


Family2_186
indication1
indication2
P
Family2_186
Chromosome 1
GBA
PGT embryo
cffDNA
Pat.FAR.seg (%)
Mat.FAR.seg (%)
Family2_186
Chromosome 2
MPV17
PGT embryo
cffDNA
Pat.FAR.seg (%)
Mat.FAR.seg (%)

c

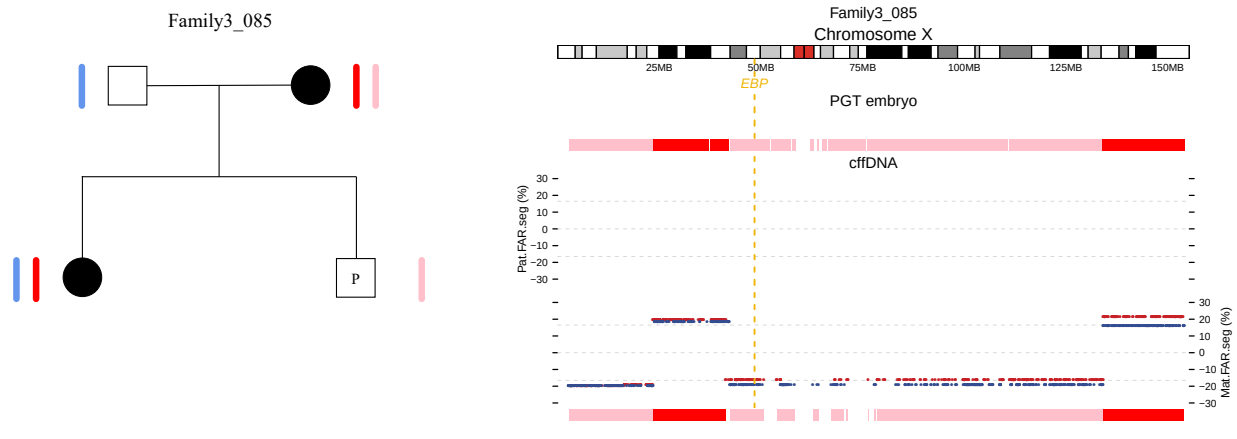

Family3_085
P
Family3_085
Chromosome X
25MB
50MB
75MB
100MB
125MB
150MB
EBP
PGT embryo
cffDNA
Pat.FAR.seg (%)
Mat.FAR.seg (%)

Figure 3

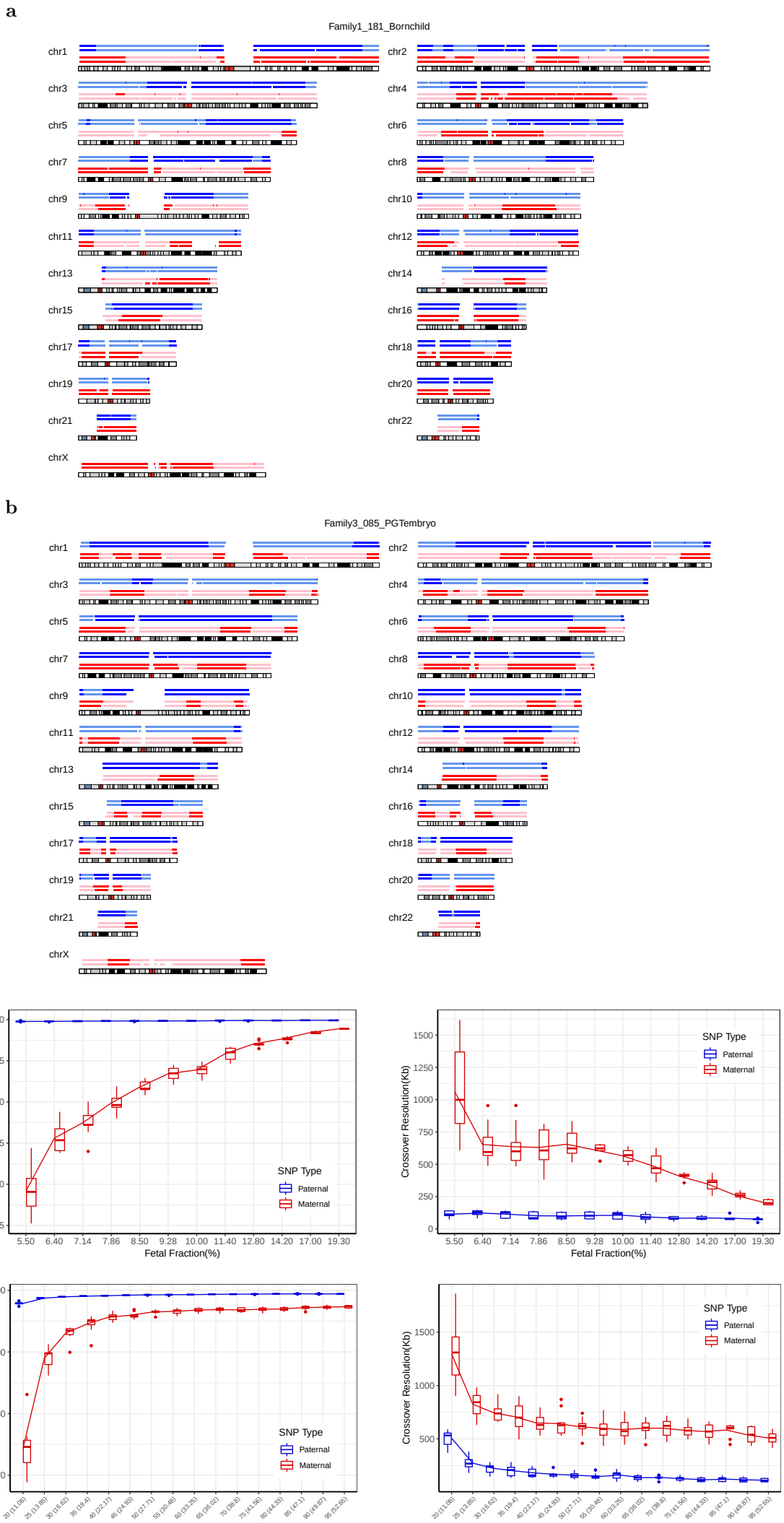
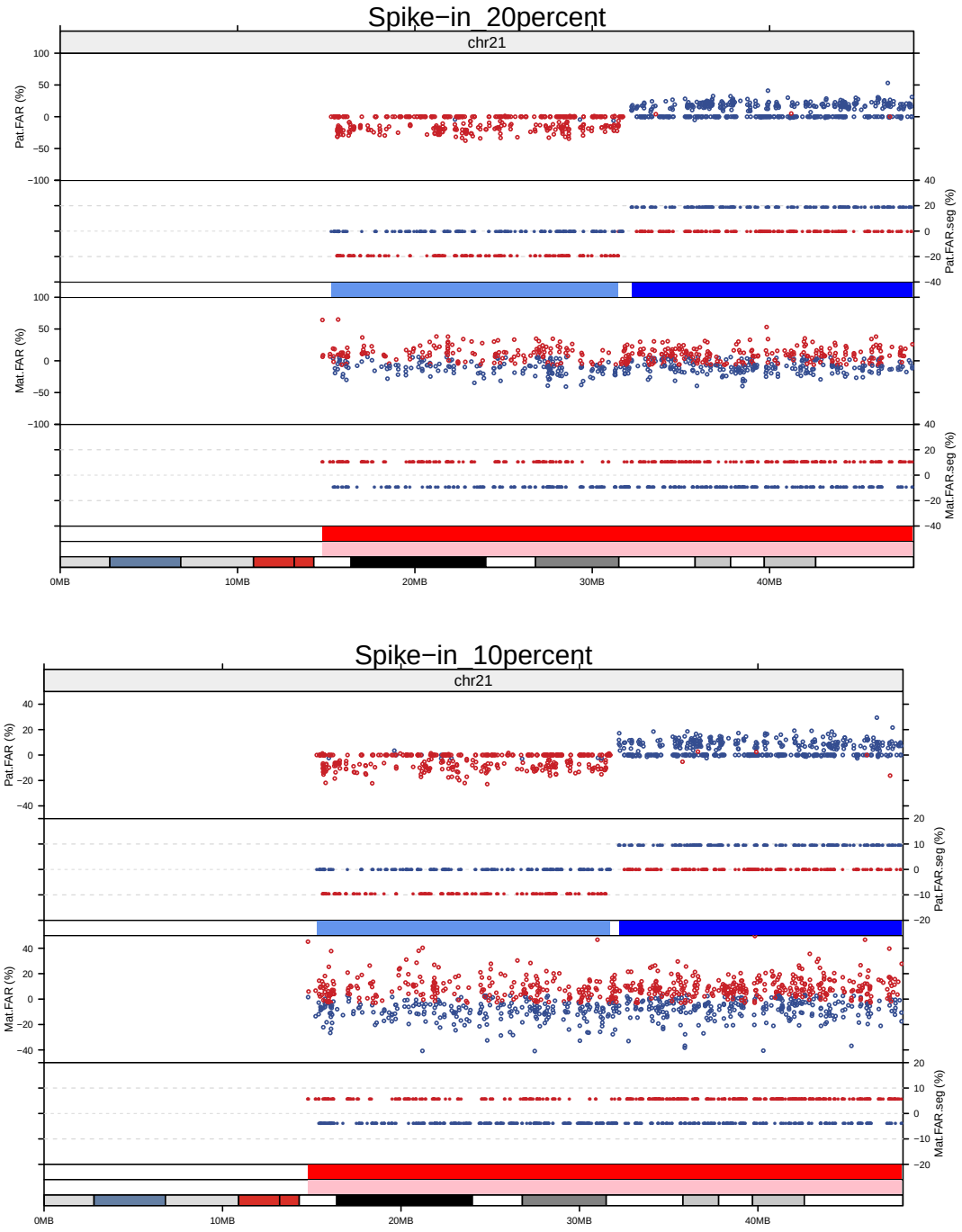
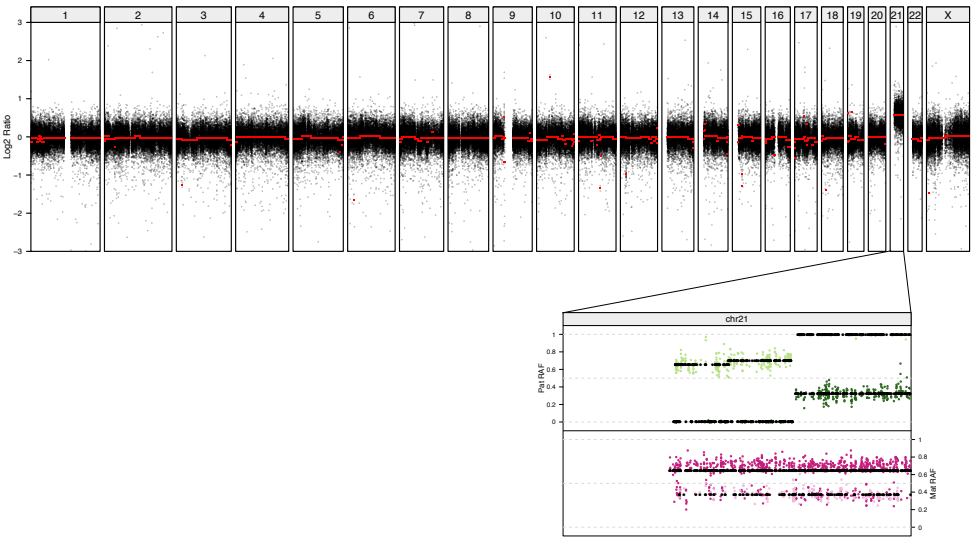

a
Family1_181_Bornchild
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chrX
b
Family3_085_PGTembryo
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chrX
c
Accuracy(%)
Fetal Fraction(%)
SNP Type
Paternal
Maternal
Crossover Resolution(Kb)
Fetal Fraction(%)
SNP Type
Paternal
Maternal
d
Accuracy(%)
Subsampling percentage (Total reads in millions)
SNP Type
Paternal
Maternal
Crossover Resolution(Kb)
Subsampling percentage (Total reads in millions)
SNP Type
Paternal
Maternal

Figure 4

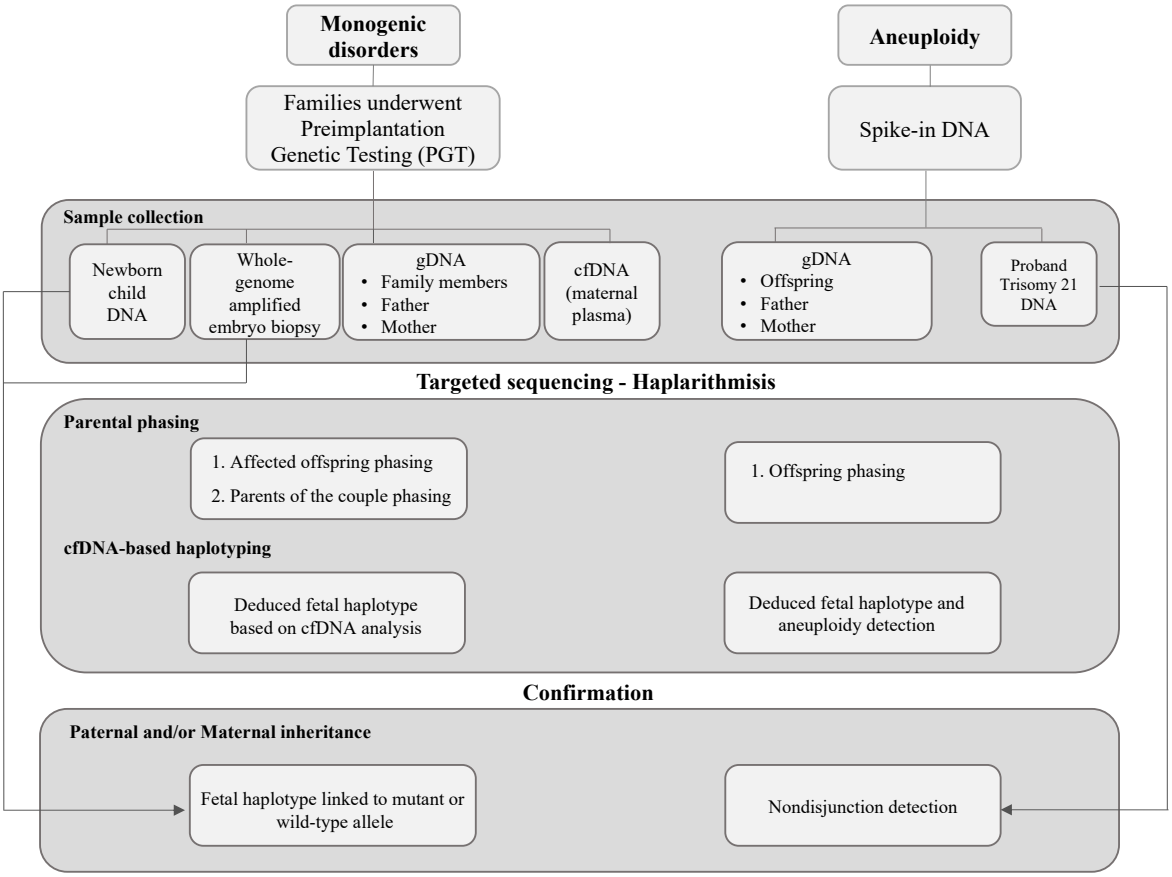
a


Spike−in_20percent
chr21
Pat.FAR (%)
Pat.FAR.seg (%)
Mat.FAR (%)
Mat.FAR.seg (%)
0MB
10MB
20MB
30MB
40MB
Spike−in_10percent
chr21
Pat.FAR (%)
Pat.FAR.seg (%)
Mat.FAR (%)
Mat.FAR.seg (%)
0MB
10MB
20MB
30MB
40MB

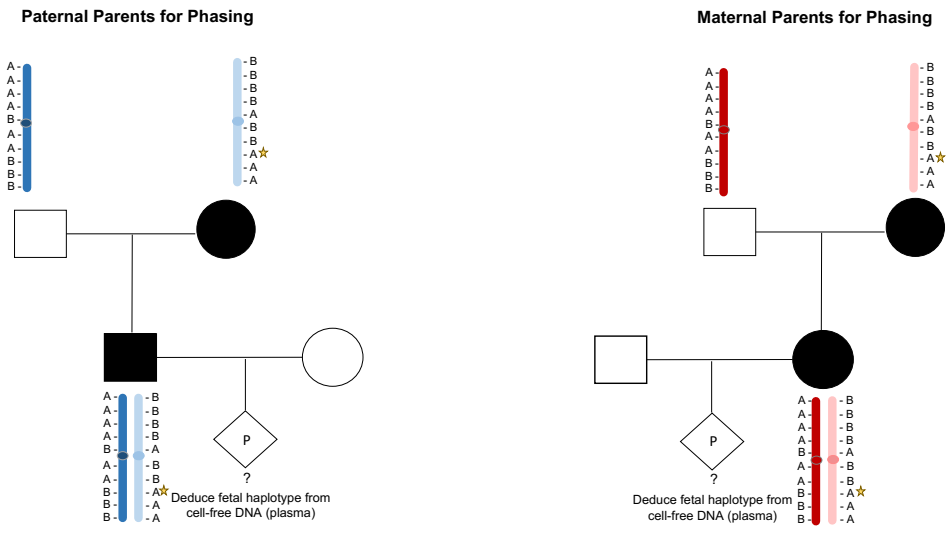
b


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Log2 Ratio
chr21
Pat.FAF
Mat.FAF

Supplementary Figure S1

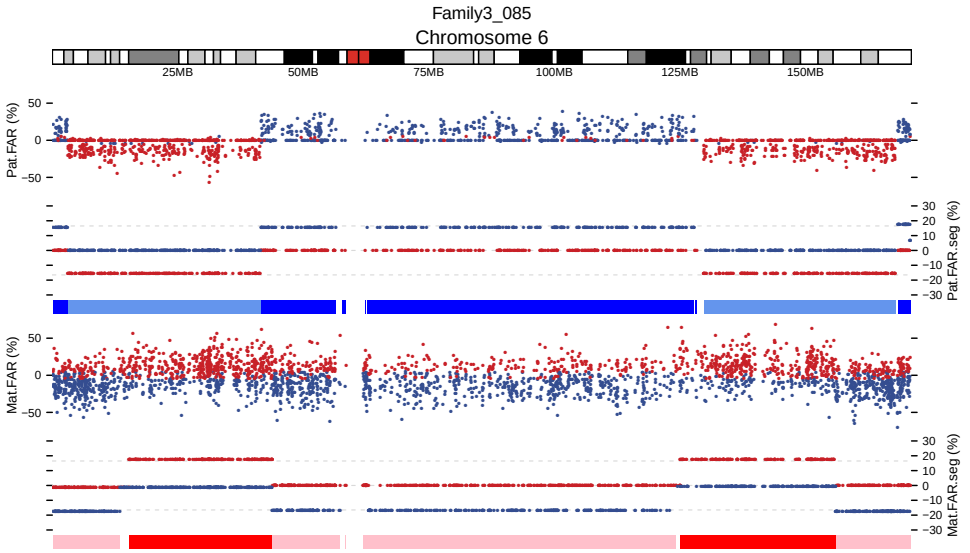


Supplementary Figure S2

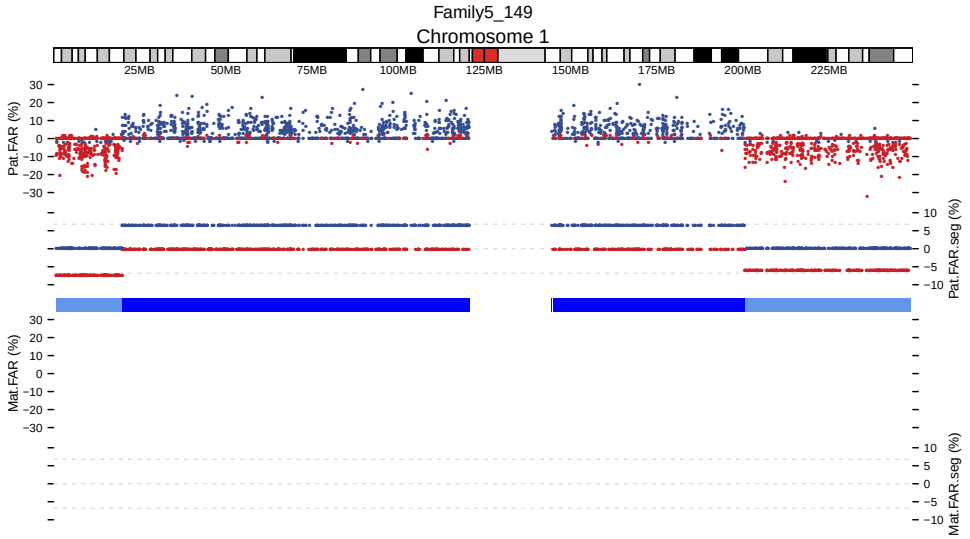


Supplementary Figure S3

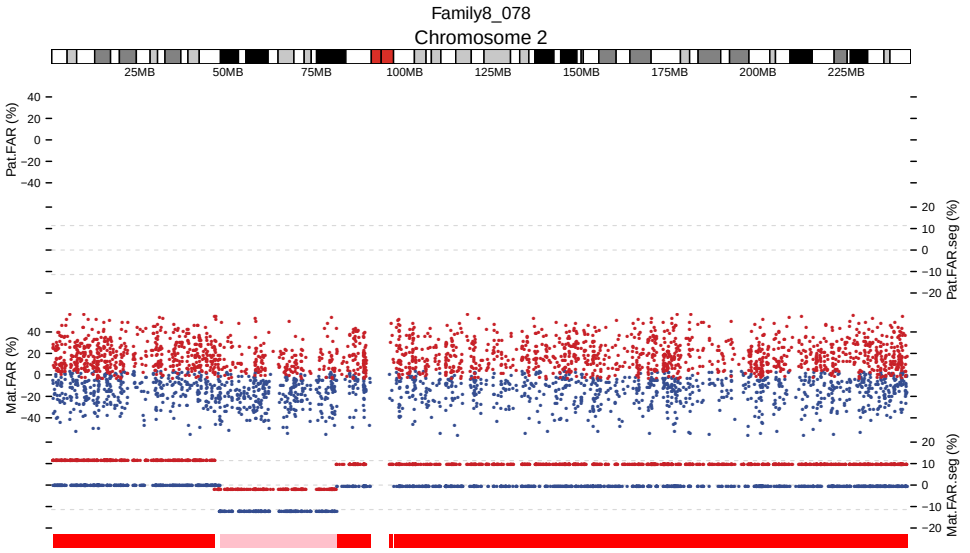
a



b

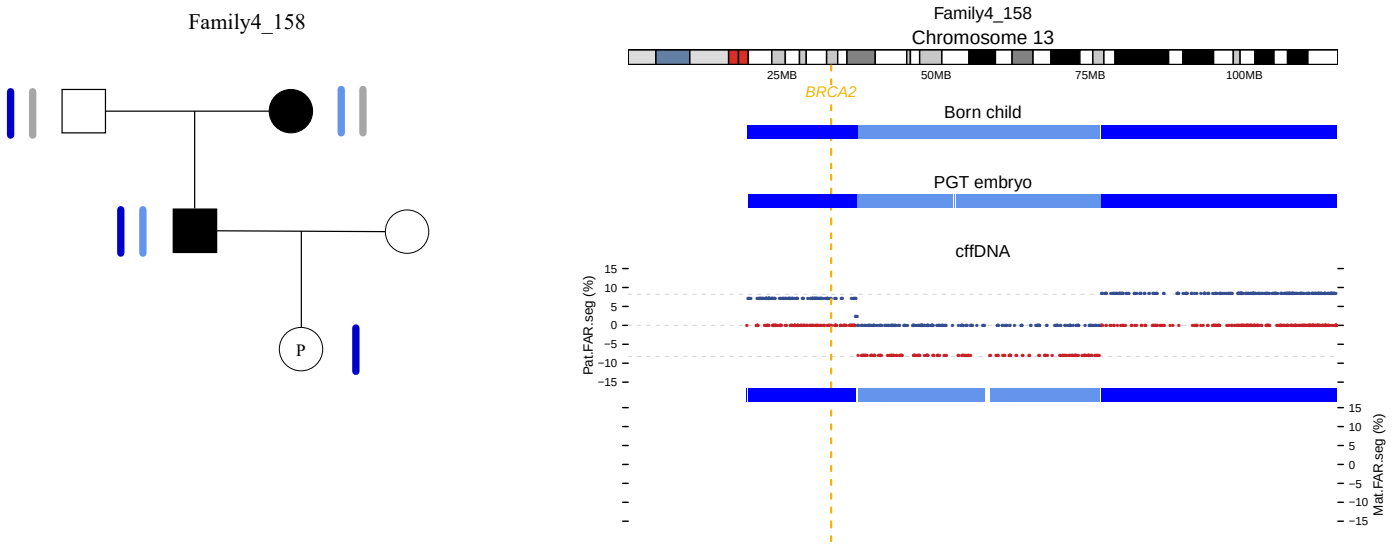


c

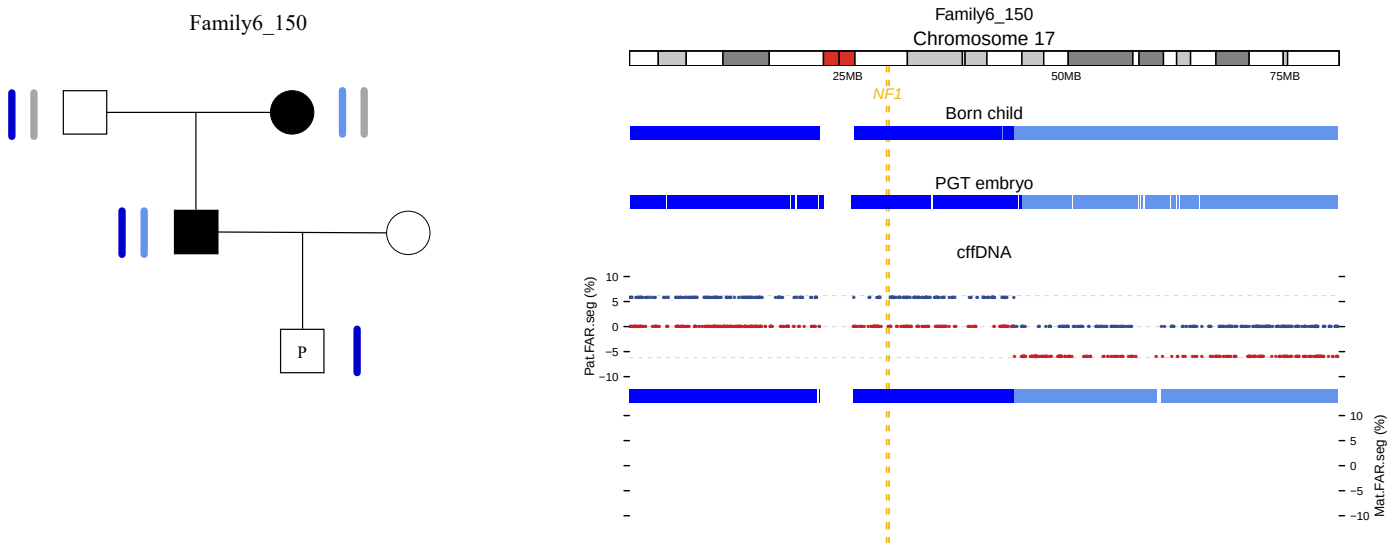


Supplementary Figure S4

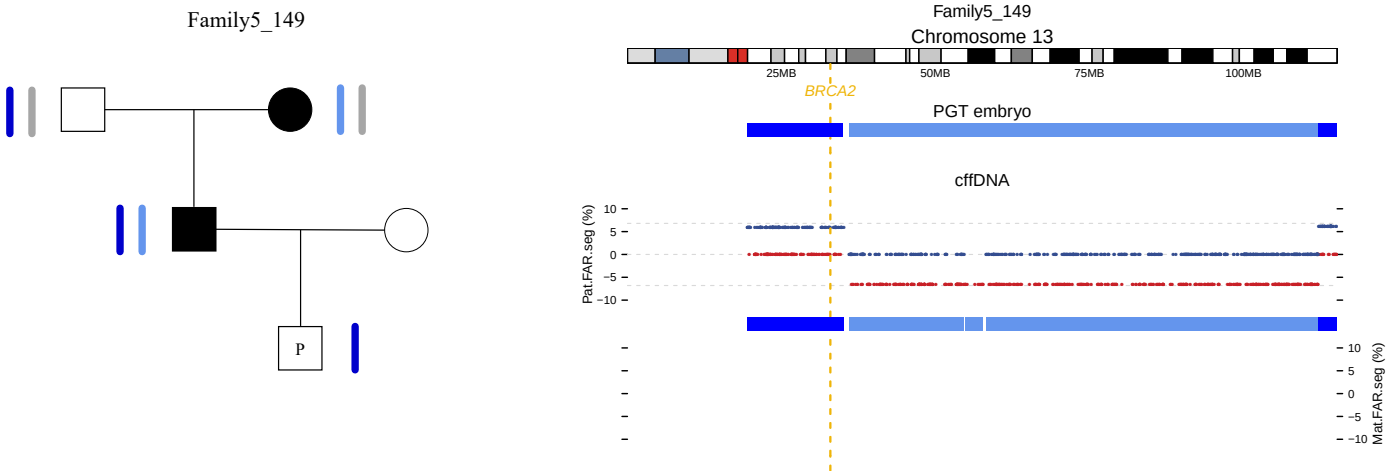
a



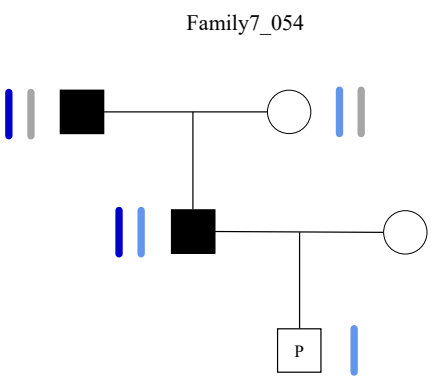
b

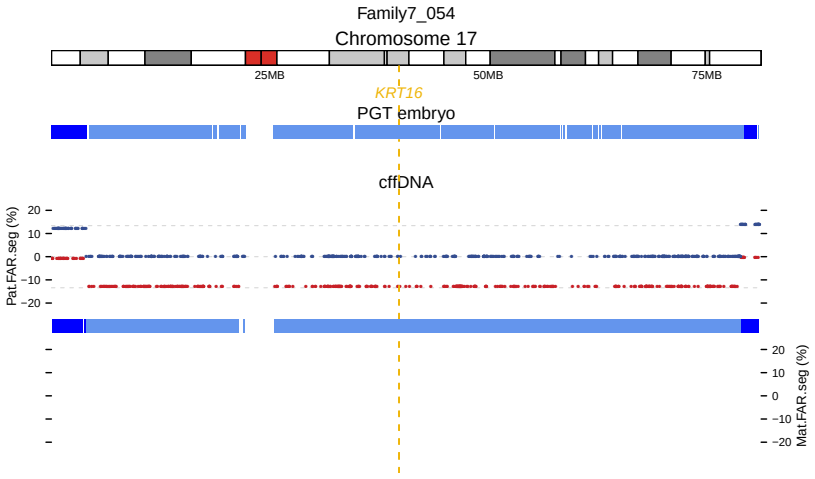


c

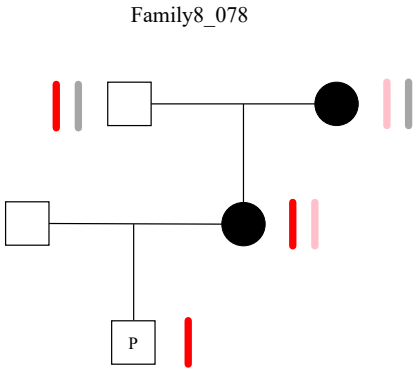


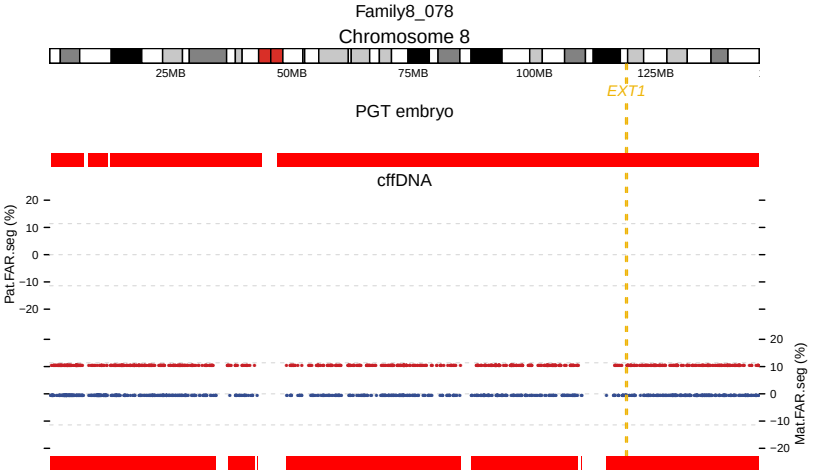
d


Family7_054
P

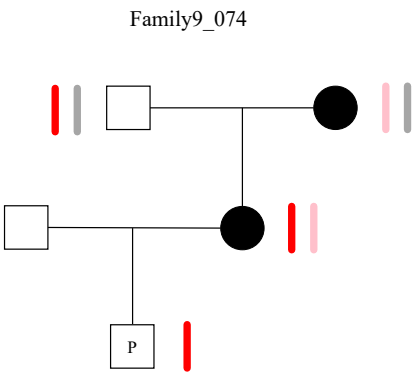

Family7_054
Chromosome 17
25MB
50MB
75MB
KRT16
PGT embryo
cffDNA
Pat.FAR.seg (%)
Mat.FAR.seg (%)

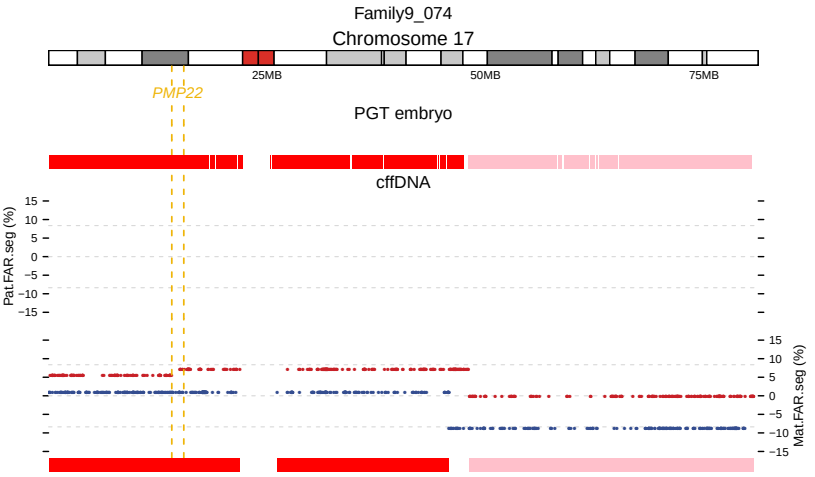
e


Family8_078
P


Family8_078
Chromosome 8
25MB
50MB
75MB
100MB
125MB
EXT1
PGT embryo
cffDNA
Pat.FAR.seg (%)
Mat.FAR.seg (%)

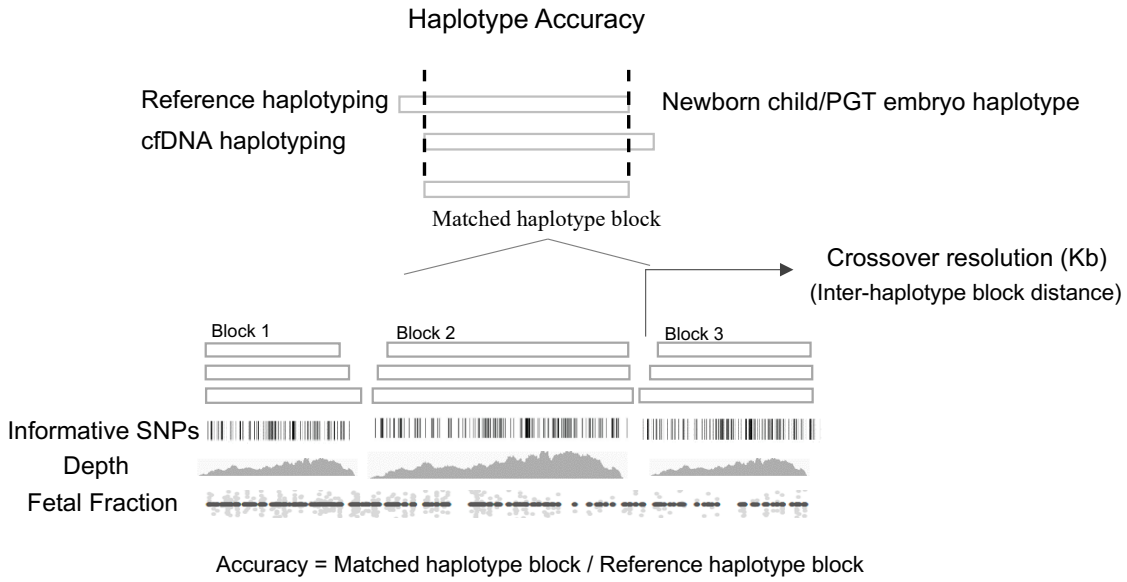
f


Family9_074
P

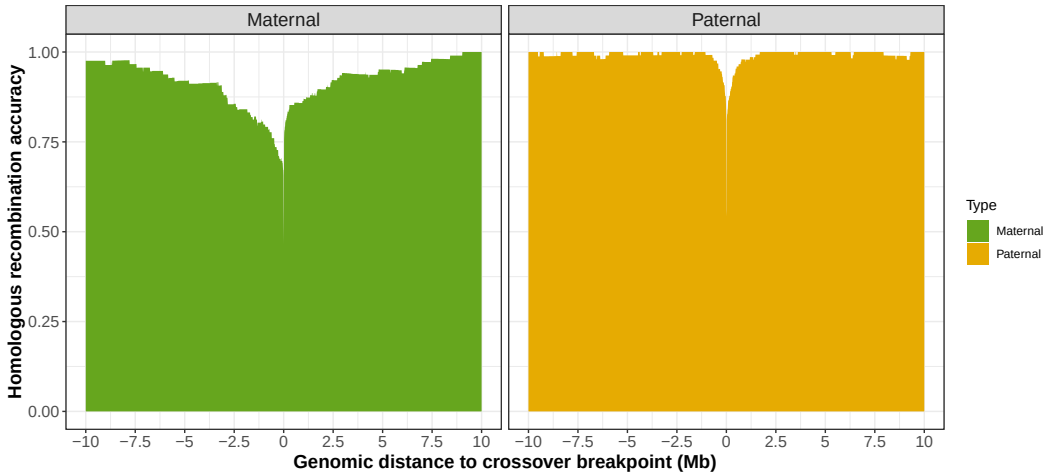

Family9_074
Chromosome 17
25MB
50MB
75MB
PMP22
PGT embryo
cffDNA
Pat.FAR.seg (%)
Mat.FAR.seg (%)

Supplementary Figure S5

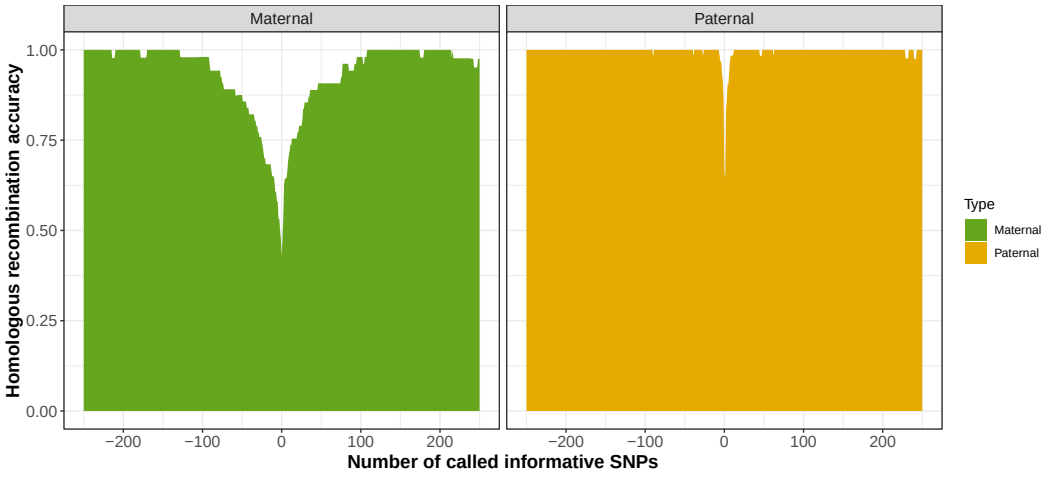
a



b

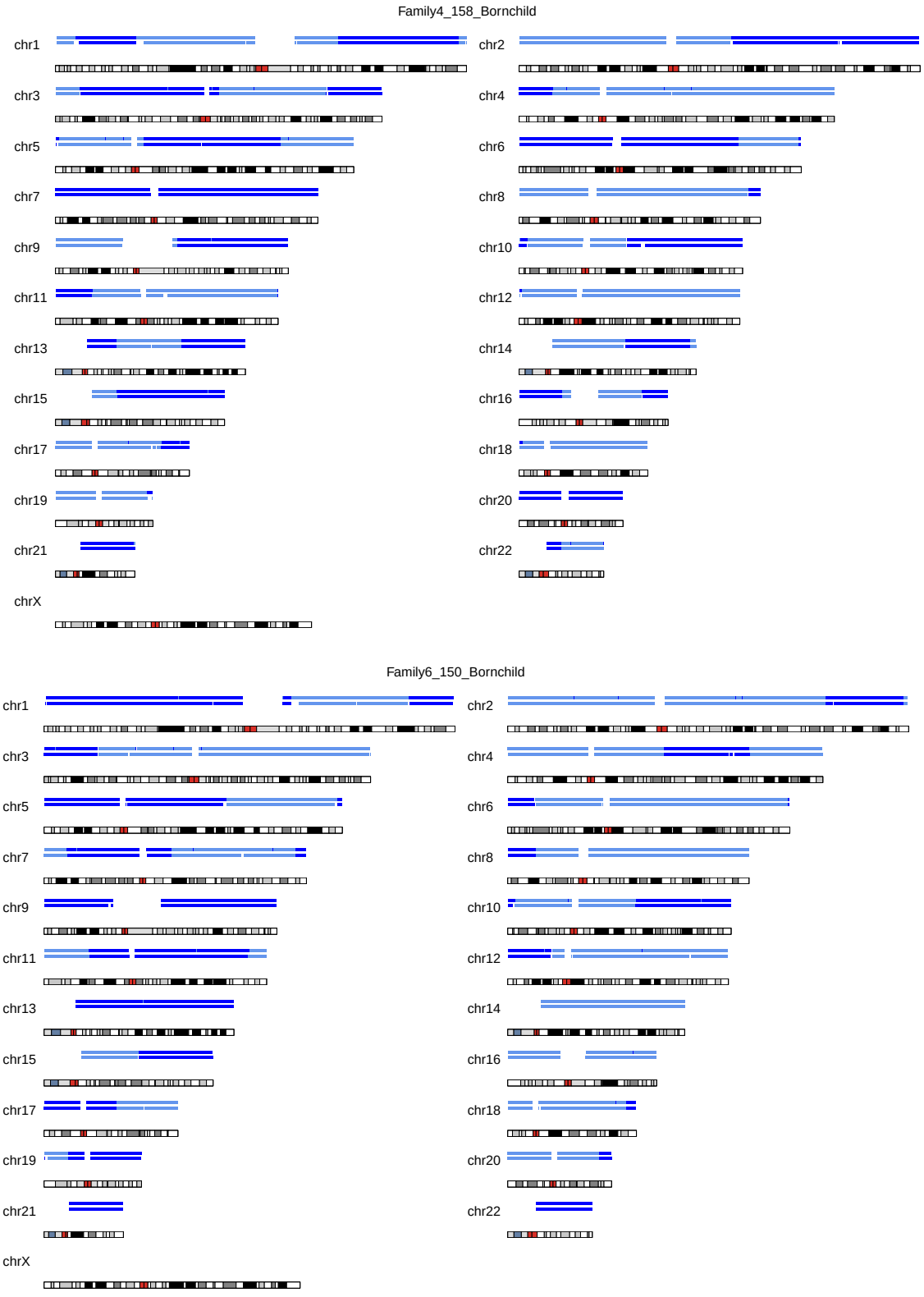


c

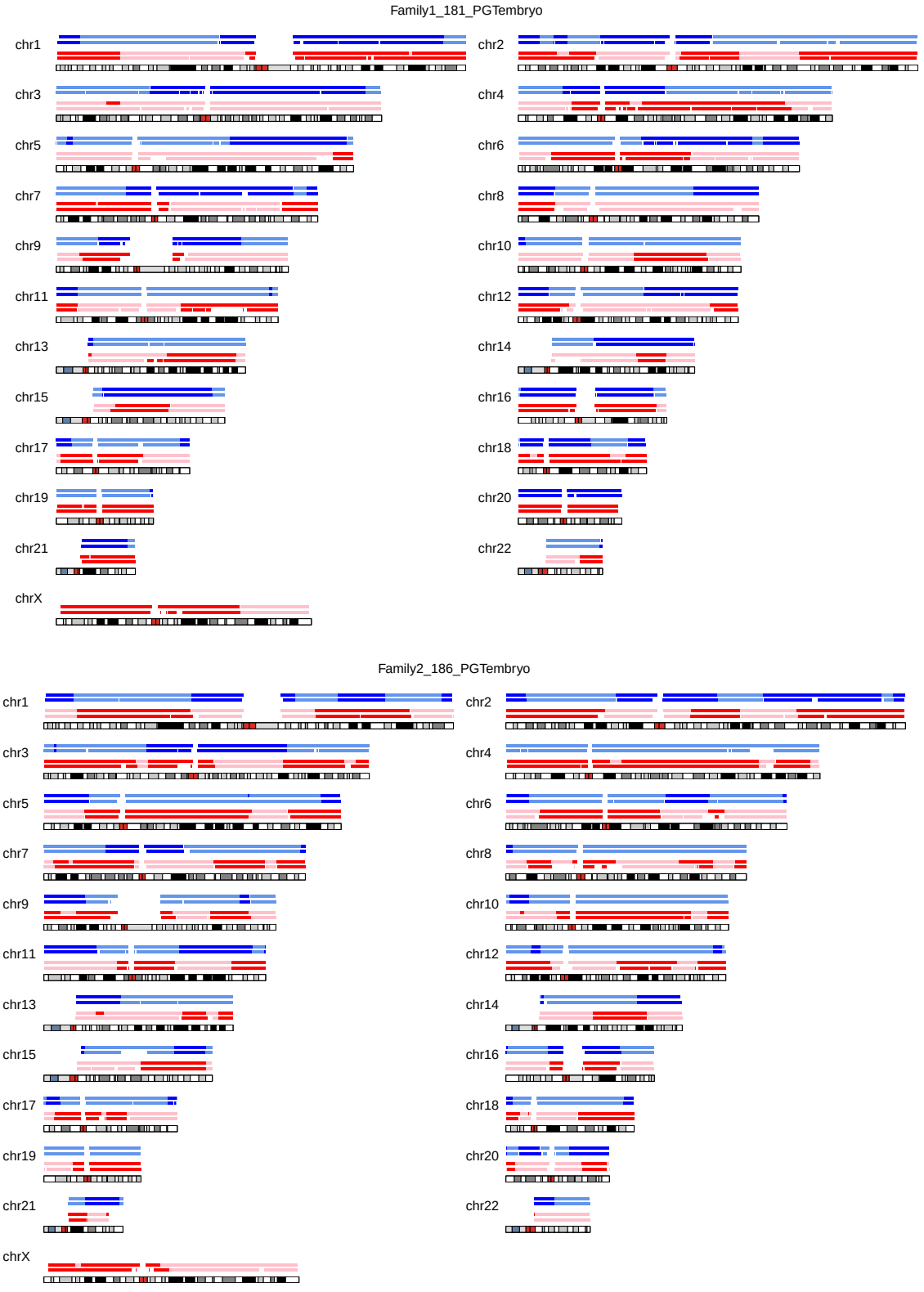


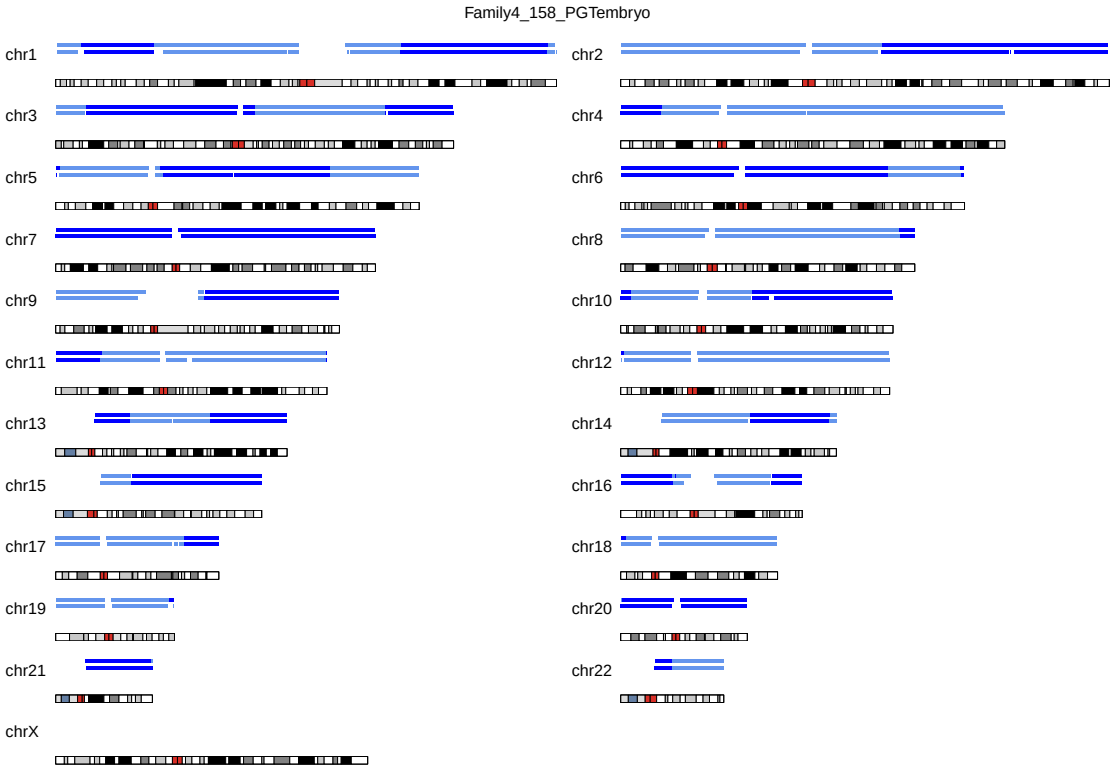
Supplementary Figure S6

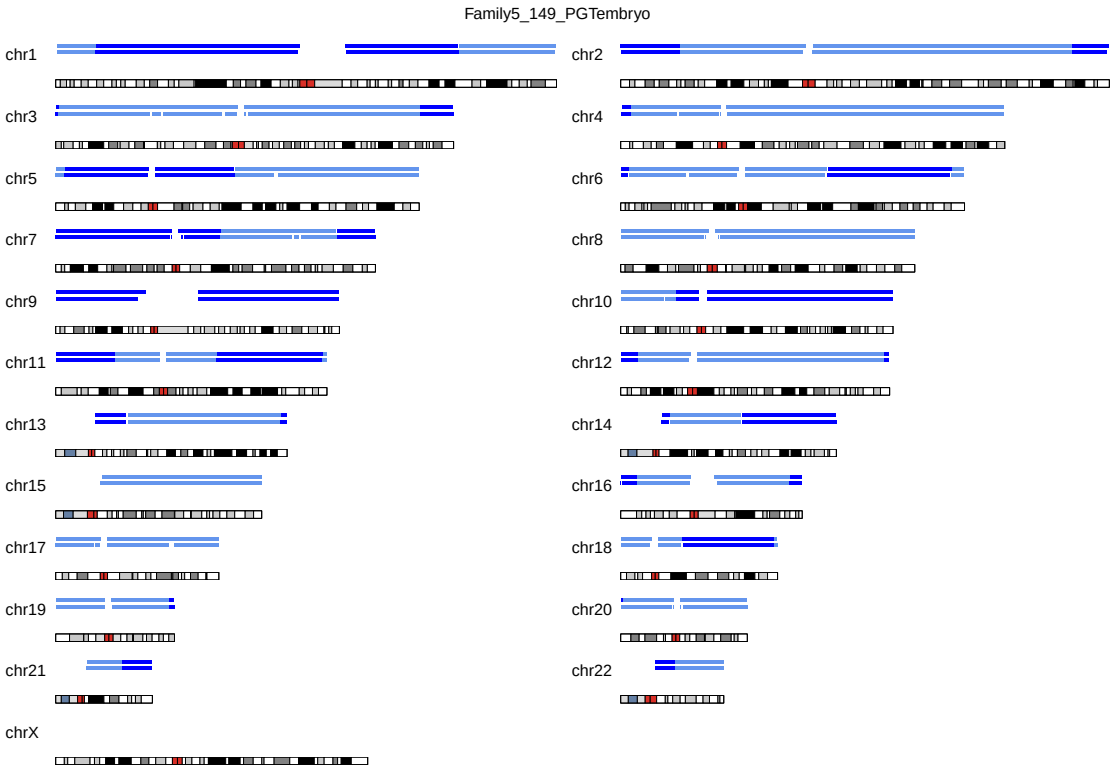
a



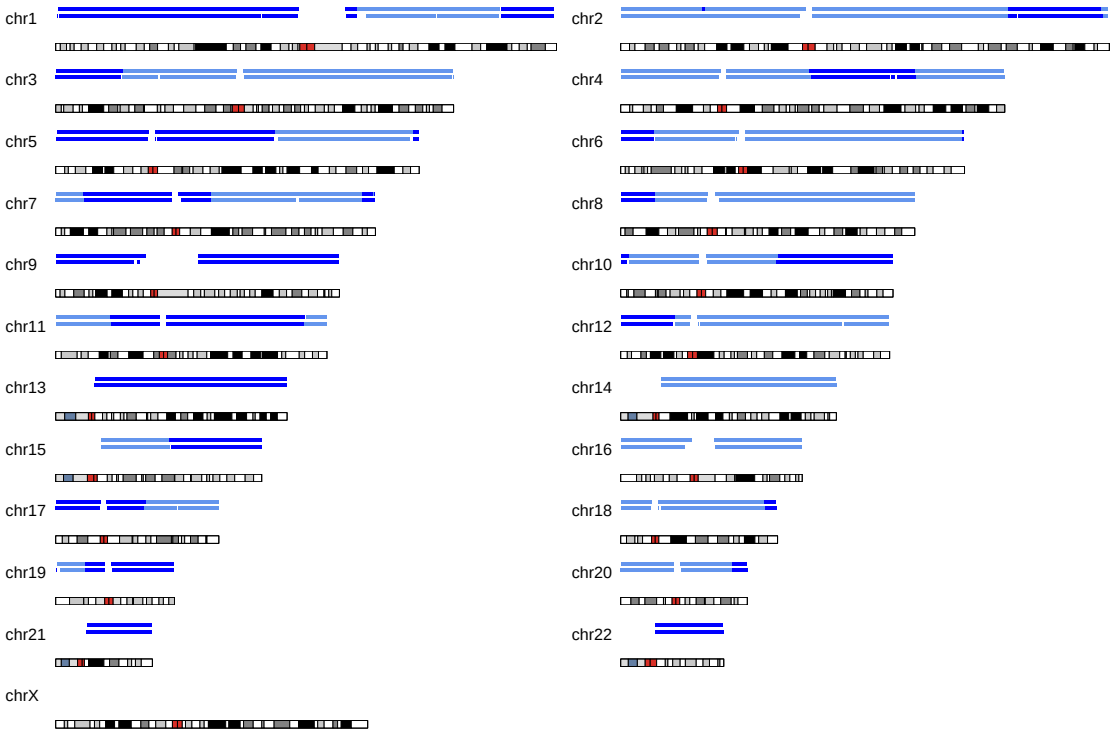
b


Family1_181_PGTembryo
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chrX
Family2_186_PGTembryo
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chrX

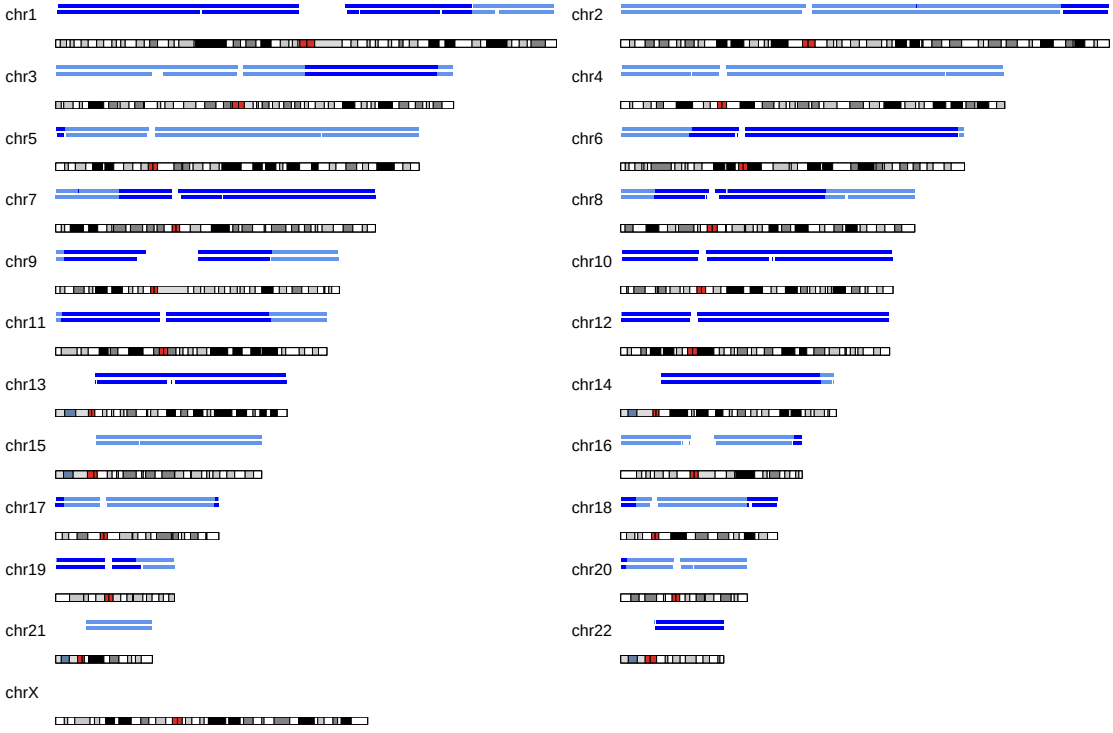

Family4_158_PGTembryo
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chrX

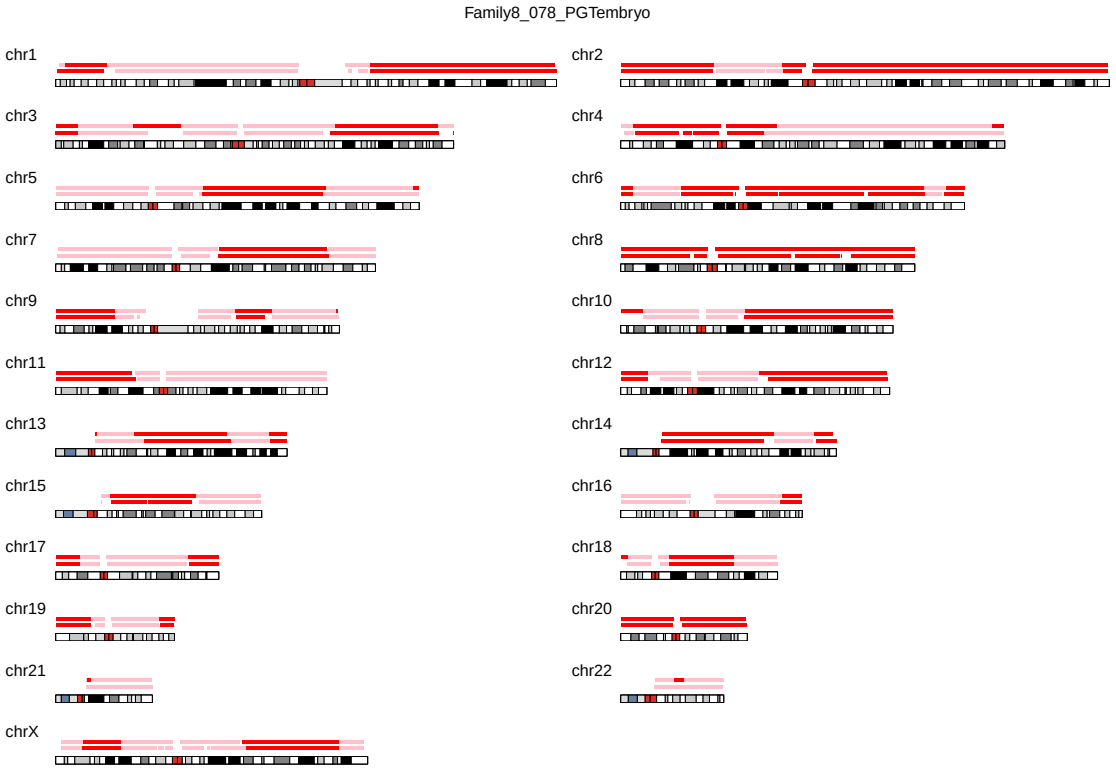

Family5_149_PGTembryo
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chrX

Family6_150_PGTembryo



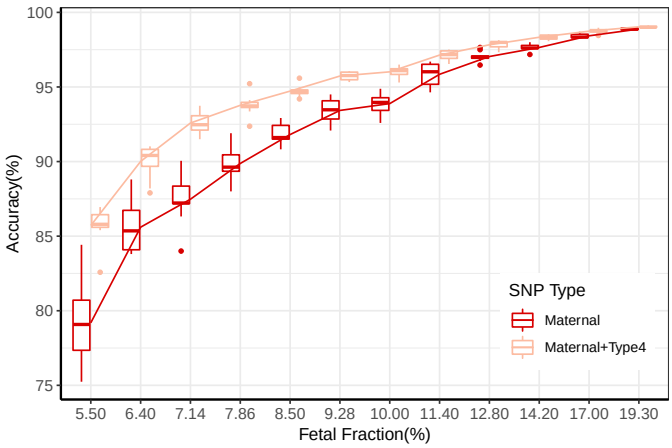
Family7_054_PGTembryo

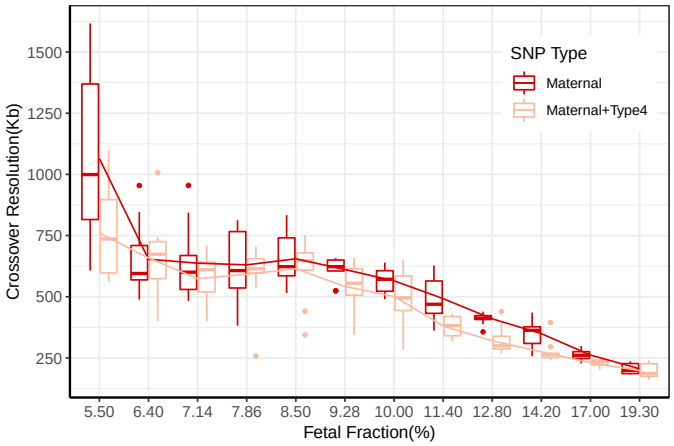



Family8_078_PGTembryo
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chrX

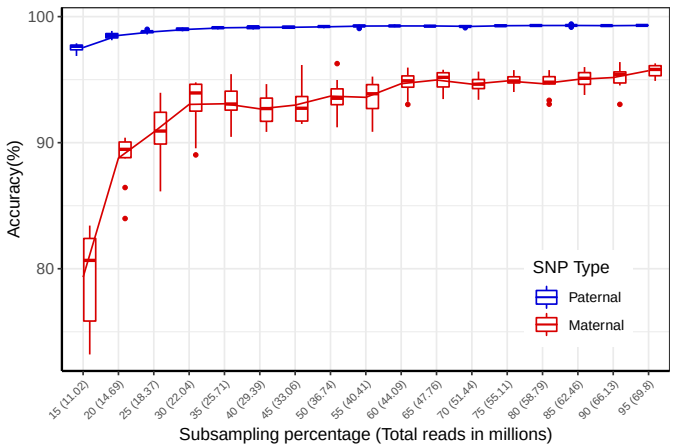

Family9_074_PGTembryo
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chrX

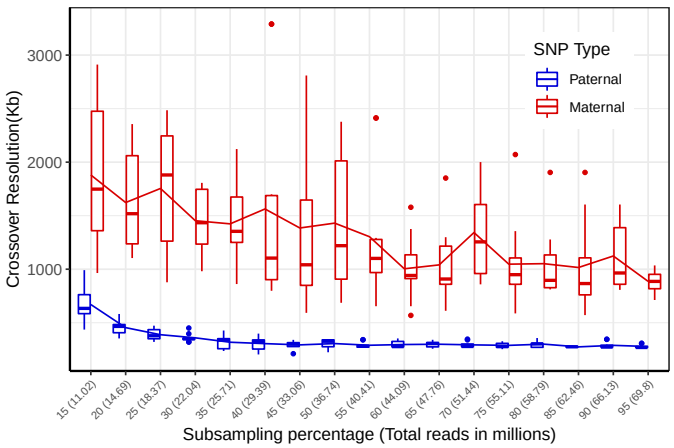
Supplementary Figure S7





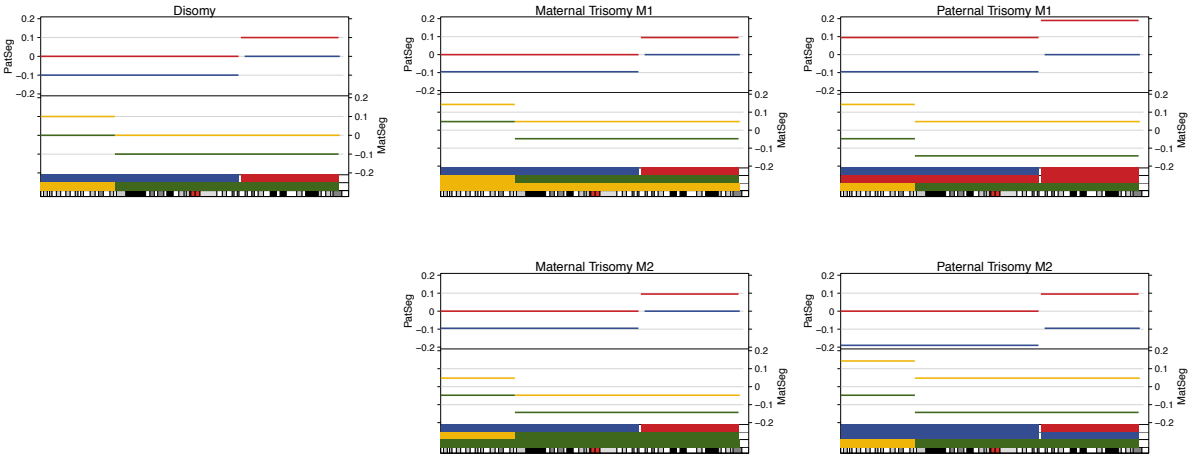
Supplementary Figure S8





Supplementary Figure S9

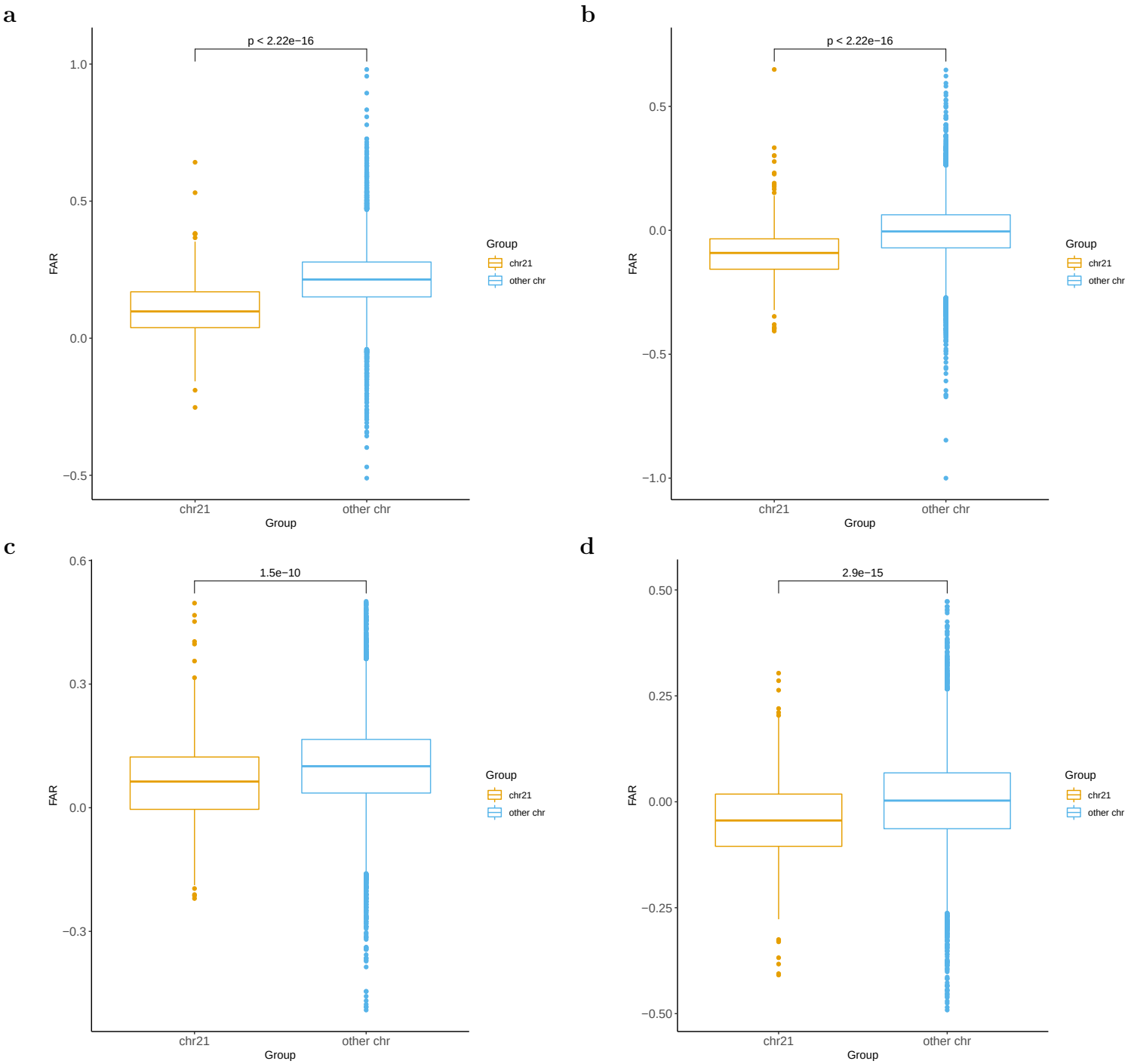
a



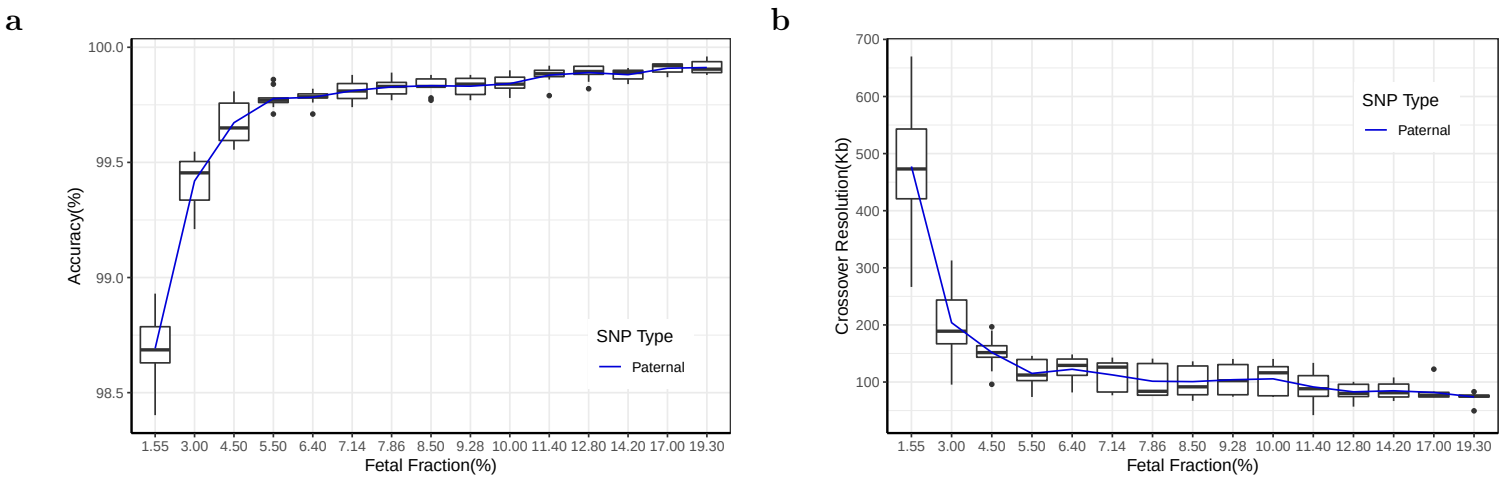
b

	Normal Disomy			Maternal Trisomy			Paternal Trisomy		
Paternal	Homolog	P1	P2	Homolog	P1	P2	Homolog	P1	P2
	H1	0	0.1	H1	0	0.0952	H1+H2	-0.0952	0.0952
	H2	-0.1	0	H2	-0.0952	0	2H1	0	0.19
							2H2	0.19	0
Maternal	Homolog	M1	M2	Homolog	M1	M2	Homolog	M1	M2
	H1	0.1	0	H1+H2	0.0476	-0.0476	H1	0.143	-0.0476
	H2	0	0.1	2H1	0.143	0.0476	H2	0.0476	-0.143
				2H2	-0.0476	-0.143			

Supplementary Figure S10

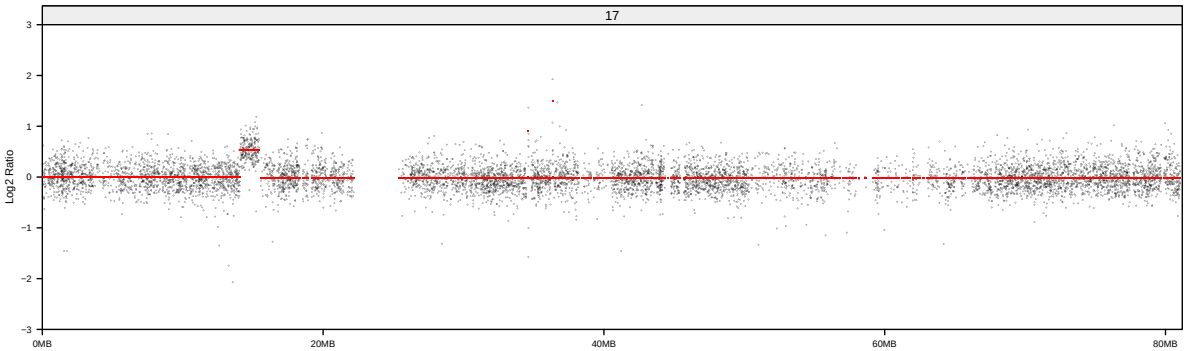


Supplementary Figure S11

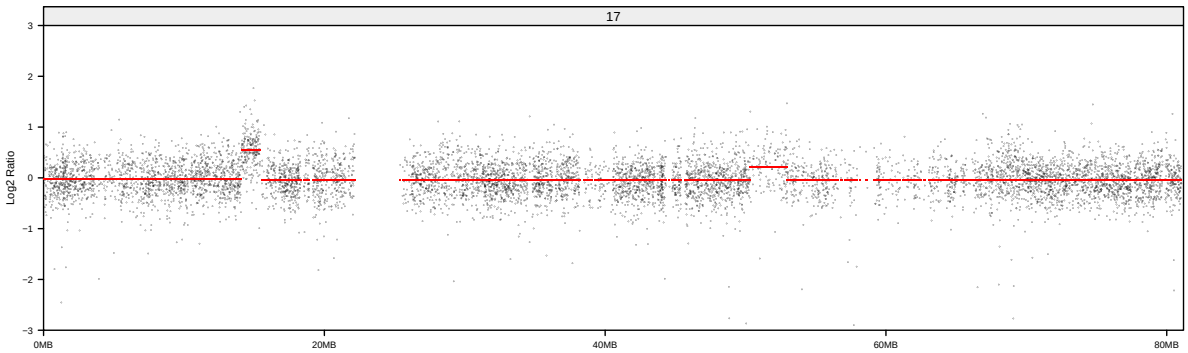


Supplementary Figure S12

a



b



Supplementary Figure S13

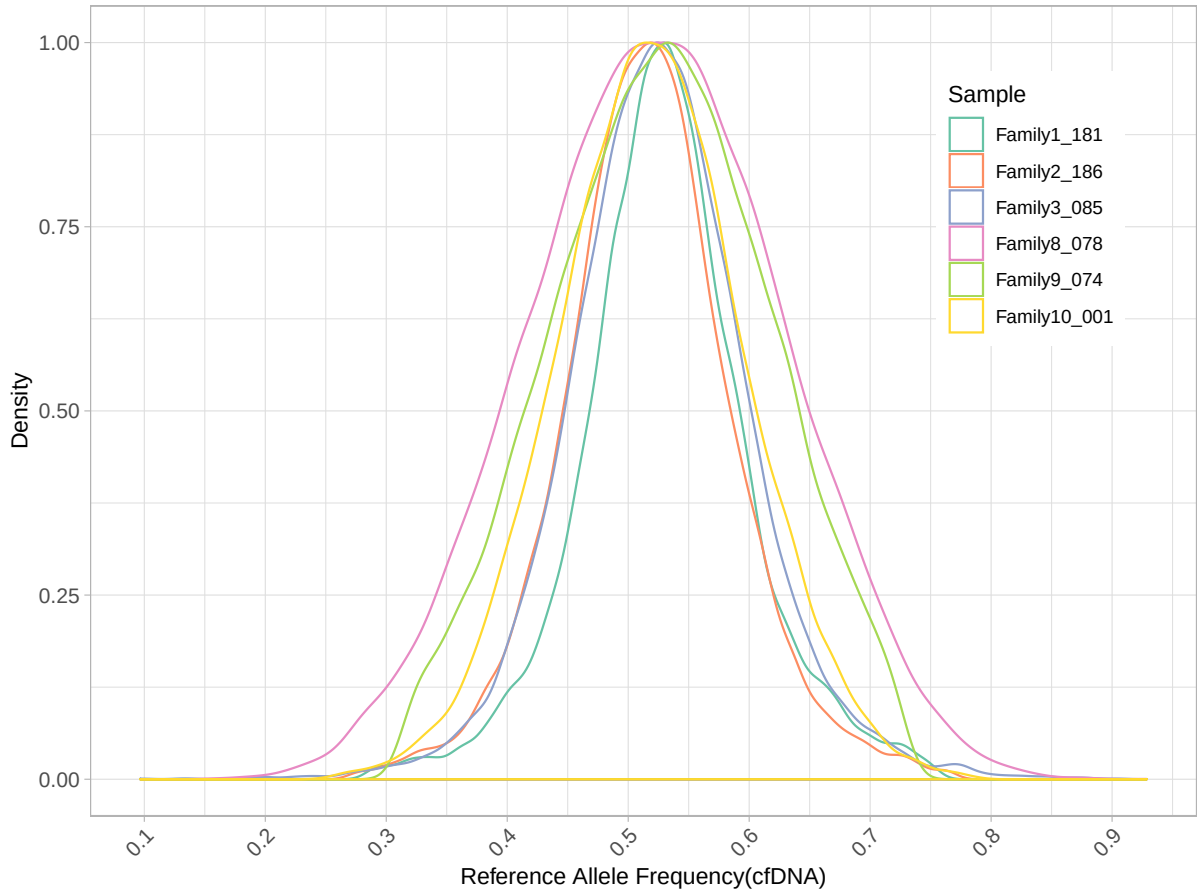


Table 1. Clinical information of all families that followed pre-implantation genetic testing (PGT)

Family	Disease	Gene	Cytogenetic Location	Genomic coordinates	MI	Mutation				Phasing	Fetal sex	Gestation age (weeks)
						Mother	Father	Other family member	Fetus			
1_181	Ceroid lipofuscinosis, neuronal, 1 (#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.556C>T)	Non carrier	Affected offspring	F	12
2_186	Gaucher disease, type 1 (#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)	Sbling homozygous c.1448T>C (p.Leu483Pro)	Non carrier	Affected offspring	F	12
	Mitochondrial DNA depletion syndrome 6 (hepatocerebral type) (#256810)	*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 mutation			
3_085	Chondrodysplasia punctata, X-linked dominant (#302960)	*EBP* (NM_006579.2)	Xp11.23	chrX:48,379,546-48,387,104	XLD	Heterozygous c.116C>G (p.Thr39Arg)	Non carrier	Sibling heterozygous c.116C>G (p.Thr39Arg)	Non carrier	Affected offspring	M	12
4_158	Breast-ovarian cancer, familial, 2 (#612555)	*BRCA2* (NM_000059.3)	13q13.1	chr13:32,889,616-32,973,808	AD	Non carrier	Heterozygous c.5645C>A (p.Ser1882*)	Paternal grandmother heterozygous c.5645C>A (p.Ser1882*)	Non carrier	Paternal parents	F	12
5_149	Breast-ovarian cancer, familial, 2 (#612555)	*BRCA2* (NM_000059.3)	13q13.1	chr13:32,889,616-32,973,808	AD	Non carrier	Heterozygous c.6275_6276delTT(p.Leu2092Profs*7)	Paternal grandmother heterozygous c.6275_6276delTT (p.Leu2092Profs*7)	Non carrier	Paternal parents	M	12
6_150	Neurofibromatosis, type 1 (#162200)	*NF1* (NM_000267.3)	17q11.2	chr17:29,421,944-29,704,694	AD	Non carrier	Heterozygous c.7096_7101delAACTTT	Paternal grandmother heterozygous c.7096_7101delAACTTT	Non carrier	Paternal parents	M	13
7_054	Pachyonychia Congenita (#167200)	*KRT16* (NM_005557.3)	17q21.2	chr17:39,766,030-39,772,151	AD	Non carrier	Heterozygous c.374A>G (p.Asn125Ser)	Paternal grandfather heterozygous c.374A>G (p.Asn125Ser)	Non carrier	Paternal parents	M	12
8_078	Exostoses multiple type 1 (#133700)	*EXT1* (NM_000127)	8q24.11	chr8:118,806,729-119,124,092	AD	Heterozygous c.2133_2151del	Non carrier	Maternal grandmother heterozygous c.2133_2151del	Non carrier	Maternal parents	M	12
9_074	Charcot-Marie-Tooth disease type 1 (#118220)	CMT1A syndrome region including *PMP22*	17p12-p11	chr17:14,097,915-15,470,903	AD	Carrier of the common ~1.5Mb duplication	Non carrier	Maternal grandmother carrier of the common ~1.5Mb duplication	Non carrier	Maternal parents	M	12

(# Mendelian Inheritance in Men); MI= mode of inheritance; AR = autosomal recessive; AD = autosomal dominant; XLD = X-linked dominant; M= male; F= female.

Metadata	
variantvalidator_version	v0.2.5
variantvalidator_hgvs_version	1.1.3
uta_schema	uta_20180821
seqrepo_db	8/21/2018

submitted_variant	validation_warnings	gene_symbol
NM_000310.3:c.451C>T		PPT1
NM_000310.3:c.566C>T		PPT1
NM_000157.3:c.1448T>C	RefSeqGene record not avail	GBA
NM_002437.4:c.293C>T		MPV17
NM_006579.2:c.116C>G		EBP
NM_000059.3:c.5645C>A		BRCA2
NM_000059.3:c.6275_6276delTT		BRCA2
NM_000267.3:c.7096_7101delAACTTT		NF1
NM_005557.3:c.374A>G		KRT16
NM_000127.2:c.2133_2151del		EXT1

Metadata	
variantvalidator_version	v0.2.5
variantvalidator_hgvs_version	1.1.3
uta_schema	uta_20180821
seqrepo_db	8/21/2018

transcript_description	HGVS_transcript_variant	HGVS_predicted_protein_conse
Homo sapiens palmitoyl-protein thi	NM_000310.3:c.451C>T	NP_000301.1(LRG_690p1):p.(Ar
Homo sapiens palmitoyl-protein thi	NM_000310.3:c.566C>T	NP_000301.1(LRG_690p1):p.(Pr
Homo sapiens glucosylceramidase b	NM_000157.3:c.1448T>C	NP_000148.2:p.(Leu483Pro)
Homo sapiens mitochondrial inner r	NM_002437.4:c.293C>T	NP_002428.1:p.(Pro98Leu)
Homo sapiens EBP, cholestenol delt	NM_006579.2:c.116C>G	NP_006570.1:p.(Thr39Arg)
Homo sapiens BRCA2 DNA repair as	NM_000059.3:c.5645C>A	NP_000050.2(LRG_293p1):p.(Se
Homo sapiens BRCA2 DNA repair as	NM_000059.3:c.6275_6276d	NP_000050.2(LRG_293p1):p.(Le
Homo sapiens neurofibromin 1 (NF1	NM_000267.3:c.7096_7101d	NP_000258.1(LRG_214p1):p.(As
Homo sapiens keratin 16 (KRT16), m	NM_005557.3:c.374A>G	NP_005548.2:p.(Asn125Ser)
Homo sapiens exostosin glycosyltra	NM_000127.2:c.2133_2151d	NP_000118.2(LRG_493p1):p.(Tr

HGVS_predicted_protein_c	HGVS_Genon	GRCh37_CHR	GRCh37_POS	GRCh37_ID
NP_000301.1:p.(R151*)	NC_000001.1	1	40555167	.
NP_000301.1:p.(P189L)	NC_000001.1	1	40546130	.
NP_000148.2:p.(L483P)	NC_000001.1	1	155205043	.
NP_002428.1:p.(P98L)	NC_000002.1	2	27535443	.
NP_006570.1:p.(T39R)	NC_000023.1	X	48382275	.
NP_000050.2:p.(S1882*)	NC_000013.1	13	32914137	.
NP_000050.2:p.(L2092Pfs*7	NC_000013.1	13	32914766	.
NP_000258.1:p.(N2366_F23	NC_000017.1	17	29670115	.
NP_005548.2:p.(N125S)	NC_000017.1	17	39768567	.
NP_000118.2:p.(W711*)	NC_000008.1	8	118812040	.

GRCh37_REF	GRCh37_ALT	HGVS_Genomic_Description_GRCh38
G	A	NC_000001.11:g.40089495G>A
G	A	NC_000001.11:g.40080458G>A
A	G	NC_000001.11:g.155235252A>G
G	A	NC_000002.12:g.27312576G>A
C	G	NC_000023.11:g.48523887C>G
C	A	NC_000013.11:g.32340000C>A
CTT	C	NC_000013.11:g.32340630_32340631del
CTAACTT	C	NC_000017.11:g.31343105_31343110del
T	C	NC_000017.11:g.41612315T>C
TCAGCGGCATGTAGCC	T	NC_000008.11:g.117799806_117799824del

GRCh38_CHR	GRCh38_POS	GRCh38_ID	GRCh38_REF	GRCh38_ALT
1	40089495	.	G	A
1	40080458	.	G	A
1	155235252	.	A	G
2	27312576	.	G	A
X	48523887	.	C	G
13	32340000	.	C	A
13	32340629	.	CTT	C
17	31343097	.	CTAACTT	C
17	41612315	.	T	C
8	117799801	.	TCAGCGGCATT	T

RefSeqGene_	genome_con	HGVS_LRG_v	HGVS_LRG_t	HGVS_RefSeq	alt_genomic_loci
		LRG_690:g.12	LRG_690t1:c.	NG_009192.1:g.12976C>T	
		LRG_690:g.22	LRG_690t1:c.	NG_009192.1:g.22013C>T	
					NW_003315906.1:g.40275A
				NG_008075.1:g.14989C>T	
				NG_007452.1	NW_004070880.2:g.76331(
		LRG_293:g.29	LRG_293t1:c.	NG_012772.3:g.29521C>A	
		LRG_293:g.30	LRG_293t1:c.	NG_012772.3:g.30151_30152del	
		LRG_214:g.25	LRG_214t1:c.	NG_009018.1:g.253129_253134del	
				NG_008301.1:g.5513A>G	
		LRG_493:g.31	LRG_493t1:c.	NG_007455.2:g.317000_317018del	

A>G;grch37;HSCHR1_2_CTG31-40275-A-G|NW_003315906.1:g.40275A>G;grch38;HSCHR1_2_CT

5C>G;grch37;HG1436_HG1432_PATCH-763316-C-G

G31-40275-A-G

Supplementary Materials and Methods

Sample collection and processing

Genomic DNAs from family members were extracted from blood following standard procedures. Briefly, 8 mL of maternal peripheral blood was collected in cell-free DNA BCT tubes (Streck, Omaha, NE, USA) or cell-free DNA collection tubes (Roche Diagnostics, Risch-Rotkreuz, Switzerland) from 10 weeks of gestation onwards. Blood was centrifuged at 1600g for 10 min at 4°C, and the plasma portion was carefully removed and centrifuged again at 16,000g for 10 min. The plasma (4 ml) and buffy coat were stored separately. DNA from the buffy coat was extracted following the "Blood and Body fluid" protocol of the QIAamp DSP DNA Blood Mini Kit (Qiagen, Hilden, Germany). cfDNA was extracted from 4 mL plasma automatically, using the Maxwell HT cfDNA kit (Promega, Wisconsin, USA) on the Hamilton Liquid Handler according to the manufacturer's recommendations.

Parental data phasing

The biallelic SNP sites from parental genotypes were extracted with lower coverage threshold of 10-fold and upper threshold of 500-fold. The parental genotypes are divided into 5 groups according to paternal and maternal allele combinations. For illustration, the biallelic sites are shown as carrying "A" or "B" allele.

Groups	Paternal genotype	Maternal genotype
Type 1	A/A	B/B
Type 2 (Paternal)	A/B	A/A
	A/B	B/B
Type 3 (Maternal)	A/A	A/B
	B/B	A/B
Type 4	A/B	A/B
Type 5	B/B	B/B

Fetal fraction estimation and error estimation

At each Type 1 SNP site, where maternal alleles are different from paternal alleles, the paternal specific allele can be identified in maternal plasma unambiguously. By counting reads that contain paternal specific alleles (*f*) and total number of reads (*N*) in such SNP sites, we estimated the fetal fraction (FF) percentage as:

$$FF = \frac{2 * f}{N}$$

Technical and biological factors contribute to inconsistent estimation of FF from site to site. Considering the variability and potential copy number changes, all paternal specific alleles were sum and divided by total reads across all Type 1 SNP loci to give FF estimation; median of FF estimation on all Type 1 SNP sites was also calculated to ensure justified estimation of FF.

Sequencing and alignment errors in maternal plasma were estimated by Type 5 SNPs. When both parents have the same genotypes (alternative alleles), the fetus is supposed to have the same genotypes, unless a *de novo* mutation happens, which is of low probability. We counted the number of reference alleles and total alleles present on Type 5 SNP sites. The ratio of reference allele numbers to total numbers were used to approximate errors (**Supplementary Table S2**).

Fetal haplotype inference based on fetal allele ratio (FAR)

The presence of fetal allele in maternal plasma was measured using the following metric – fetal allele ratio (FAR)

$$FAR_i = \begin{cases} \frac{2 * f_i}{N_i}, & SNP\ site\ i\ is\ Type\ 2\ SNP \\ \frac{MH1_i - MH2_i}{MH1_i + MH2_i}, & SNP\ site\ i\ is\ Type\ 3\ SNP \end{cases}$$

FAR calculations combined with subcategory classifications estimate fetal genome-wide

haplotype. FAR value changes in subcategory indicate the homolog changes in inheritance, and thus perform phasing of recombinant haplotypes. In paternal transmission, we flipped the FAR value of P1 category by using 0-FAR for visualization purpose. Assuming 100-fold coverage and 10% FF, when paternal homolog H1 is inherited, FAR of P1 SNPs will be 0 and FAR of P2 will be 10%, as shown in the example **Figure 1b**, and when paternal homolog H2 is inherited, FAR of P1 SNPs will be -10% and FAR of P2 will be 0. In maternal transmission, FAR of M1 SNPs will be 10% and FAR of M2 will be 0 when maternal homolog H1 is inherited, and FAR of M1 SNPs will be 0 and FAR of M2 will be -10% when maternal homolog H2 is inherited. FAR values of subcategories (P1 and P2 or M1 and M1) hence infer haplotype results reciprocally.

While FAR provides information on the allele dosage of the fetus and site-by-site genotype of informative SNPs could be predicted, there is high variability in practical estimation due to technical bias, biological features of plasma DNA, and dominant maternal DNA in maternal plasma. To overcome the noises in site-by-site FAR estimations, we employed circular binary segmentation[1] (CBS) to estimate fetal allele ratio in contiguous regions and to identify likely recombination sites. Briefly, $FAR_1, \dots, FAR_i$ raw values were calculated for Type 2 or Type 3 sub-categorical SNPs, and the data was smoothed by adjusting outlier values, resulting in $\widetilde{FAR_1}, \dots, \widetilde{FAR_i}$. The likelihood ratio test was then performed on smoothed data to test the null hypothesis of no change-point in the data. If the test statistic exceeds threshold and reject the null hypothesis, the change points were identified. By detecting these change points in fetal allele ratio values, which correspond to homolog inheritance changes, we deduce the possible genomic locations of switchovers. The haplotype blocks derived from two subcategories were screened after the segmentation to avoid long regions that contains no informative SNPs. Within one block, if the genomic distance between two adjacent

informative SNPs were greater than 3Mb, the block was divided into two blocks with the two end positions (end of block 1 and start of block 2) corresponding to the two SNPs genomic positions. The segmental mean FAR estimates from subcategory P1 and P2 or M1 and M2 were examined to check whether both subcategories segmentation reach the same conclusion. We concluded that paternal H1 was inherited only when $\overline{FAR_{P1}}$ was in the range $(-0.25 * FF, 0.25 * FF)$ and $\overline{FAR_{P2}}$ was in the range $(0.75 * FF, 1.25 * FF)$, and paternal H2 was inherited when when $\overline{FAR_{P1}}$ was within $(-1.25 * FF, -0.75 * FF)$ and $\overline{FAR_{P2}}$ was within $(-0.25 * FF, 0.25 * FF)$. The same FAR value thresholds applied to maternal H1 and H2 inheritance. Fetal allele ratio value that fell out of threshold range was defined as inconclusive result.

Newborn haplotype

The newborn child's sequencing data was analyzed in conjunction with the parental genotypes of the family. Genotypes of the family trio were called together with HaplotypeCaller. We matched the child genotypes with phased parental genomes that were resolved using affected offspring or parents of the couple. Then parental homologous that defined by phasing with relatives were assigned to matched child genotypes site by site. SNPs with unambiguously assigned homolog 1 or homolog 2 were used to infer child haplotype. In cases where parental haplotypes were defined by the affected offspring, there might be small recombinations shown on child haplotype due to the crossovers from the sibling and the child happen in close genomic locations.

Accuracy compared with born child haplotypes were measured in haplotype block length and informative SNPs inference. Haplotype accuracy was calculated as the percentage of matched haplotype blocks length divided by the total length of haplotypes obtained in born child

haplotypes. Informative SNPs inference accuracy was measured by matching correctly assigned SNP genotypes in cfDNA using haplotypes information to born child genotypes.

Reference bias correction

We observed bias towards reference allele at heterozygous sites in cfDNA samples. **Supplementary Figure S11** showed the reference allele frequency of heterozygous sites in multiple cfDNA samples, where both maternal and fetal SNPs were assumed to be heterozygous with haplotype information available. For M1 subcategory, when the fetus inherits maternal H2 from mother, the fetus genotypes are heterozygous in the SNP sites within the haplotype block. For M2 subcategory, the fetus genotypes are heterozygous when maternal H1 is inherited from the mother. By computing reference allele frequency of corresponding heterozygous sites for M1 and M2 separately, the allele counts displayed bias towards reference allele. Though there might be less bias towards reference allele at SNPs where the genotype of the mother is heterozygous and the fetus is homozygous, the dominant maternal alleles skews the allele ratio. Maternal haplotyping inference can be distorted due to such bias. We performed bias correction by imposing a correction factor to reference and alternative alleles count in plasma samples to reduce bias. We used an arbitrary correction factor of 0.525 to estimate putative haplotype inheritance. Additionally, from estimated haplotype we sampled SNPs to approximate true reference allele bias per chromosome and per subcategory using maximum likelihood estimation. The computed correction factors were applied to the reference and alternative allele counts to adjust value of FAR for each SNP.

Fetal fraction simulation and cfDNA downsampling

We mixed different proportions of reads computationally from the trisomy family where the mother and the child genomic DNA sequencing were available. Samtools v1.3.1[2] were used

to randomly subsample reads from bam files and in total about 70 million reads (~85x on target coverage) were sampled from the mother and the child. By mixing all reads from mother and 4% of reads from the child, FF of 5.5% was simulated. To evaluate lower FF limit for paternal inheritance haplotyping, all reads from mother and less than 4% of reads from the child were mixed, though the total number of reads decreased slightly in this case. FF were estimated using mixture proportions. For each sampling proportion pairs, we repeated for 10 times with different sampling seeds. The child haplotype was phased directly from family trio analysis using GATK PhaseByTransmission. Haplotyping results from *in silico* simulation samples were compared to the reference child haplotype to measure genome-wide accuracy and crossover gaps. Chromosome 21 were excluded from accuracy benchmarking.

We downsampled the cfDNA sample to different proportions using samtools. Each subsampling proportion was repeated for 10 times using different seeds. Downsampling was conducted until only 11 million reads were left. We again measured the accuracies against PGT embryo haplotying result and crossover resolutions for all subsampling levels.

Integration of Type 4 SNPs in cfDNA haplotyping analysis

In cases where offspring (sibling of the born child) genotype data were available, we used Type 4 SNPs to improve haplotyping accuracy. Type 4 SNPs are defined as heterozygous SNPs for both parents and these SNPs can be phased (i.e. the genotype of sibling on these SNPs sites are homozygous). We first inferred paternal haplotyping with Type 2 SNPs. Using paternal haplotype information, we could deduce the allele that fetus inherited from a Type 4 SNP when the site was within resolved haplotype block, and thus the site became unambiguous. We then regarded the Type 4 site as a homozygous site and turned the site into

a Type 3 SNP site, as the paternal genotype was viewed as homozygous and maternal genotype is heterozygous. We further integrated unambiguous Type 4 SNPs into Type 3 SNPs and performed same analysis as when we only use Type 3 SNPs.

Aneuploidy detection

Genomic DNA from all the family members (father, mother, unaffected sibling, and affected proband) was extracted following standard procedures. Parental genotype phasing was done using an unaffected offspring from the family, and haplotype inheritance was deduced from the spike-in DNA samples. Chromosomal abnormalities result in FAR value deviation from the expected FF. The shifting pattern is associated with FF. In maternal meiotic trisomy, the overrepresented allele leads maternal FAR value to $\pm\frac{1}{2}FF/(1+\frac{1}{2}FF)$, while in mitotic trisomy, M1 FAR drifts to $\pm\frac{3}{2}FF/(1+\frac{1}{2}FF)$ and M2 FAR to $\pm\frac{1}{2}FF/(1+\frac{1}{2}FF)$. Paternal FAR value deviates from normal FF by $\pm FF^2/(2+FF)$, which is of very low amount to be distinguished from normal disomy cases.

Aneuploidy confirmation in the trisomy child and maternal copy number

As the trisomy child is a male, we used fifteen normal male gDNA samples from our clinical dataset to build a reference set. The normalized coverage per target was calculated by total coverage within the target divided by the target region size. We calculated GC content of each target and used Loess regression to correct GC bias. A further step of normalization was carried out by dividing the GC corrected median target coverage count. The reference set summarized the average normalized count and standard deviation across the 15 samples. Log2 ratio per target was computed using normalized count from the trisomy child gDNA sample against the corresponding normalized count in the reference set. We used CBS

segmentation on log2 ratios to detect copy number change as previously described. To infer homologous recombination sites and origin of aneuploidy on chromosome 21, we analyzed the family data using the child reference allele frequency count. Maternal copy number detection of family 9_074 used same approach as the aneuploidy detection, except that for cfDNA maternal copy number, a reference was created by 8 cfDNA from other families.

Supplementary References

1. Shen, J. J. & Zhang, N. R. Change-point model on nonhomogeneous Poisson processes with application in copy number profiling by next-generation DNA sequencing. *Ann. Appl. Stat.* **6**, 476–496 (2012).
2. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

Supplementary Figure Legends

Supplementary Figure S1. Study workflow.

Supplementary Figure S2. **Scheme of parental genotype phasing via parents of the couple**. **a**, Paternal haplotype derived from paternal parents with the mutation inherited from grandmother. **b**, Maternal haplotype derived from maternal parents with the mutation inherited from grandmother.

Supplementary Figure S3. cffDNA haplotyping profiles with different phasing options. Chromosome-specific cffDNA haplotyping plots are depicted for three phasing options. In **a**, a family with an affected child were used to deduce paternal (upper panel) and maternal haplotype (lower panel) inheritance. The first track shows raw fetal allele ratio (FAR) values, and each dot represents an informative SNP, with red and blue indicating paternal P1 and P2 SNPs subcategories, respectively. We flipped FAR for P1 subcategory to negative values for better visualization. The same red and blue color pattern was also applied to distinguish maternal M1 and M2 SNPs subcategories. The second track corresponds to FAR segmentation, where 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. The third track is the deduced fetal haplotypes inherited from father or mother. Dark and light blue denotes inheritance of homologue 1 (H1) and homologue 2 (H2) from the father, respectively. In the same way, dark and light red denotes inheritance of homologue 1 (H1) and homologue 2 (H2) from the mother, respectively. Switch between H1 and H2 denotes homologues recombination sites, which are also corroborated by the segmented FAR profiles. Centromere and regions with inconclusive

haplotype results are shown as white gap in the final haplotype result. **b**, Parents of the father are available for phasing and only paternal haplotype inheritance is shown. **c**, Parents of the mother are available for phasing and only maternal haplotype inheritance is shown.

Supplementary Figure S4. cffDNA haplotyping for monogenic disorders with phasing from parents of the couple. **a-d**, Families with parents of the father available for phasing. **a-b**, haplotypes inferred from cfDNA were compared to both neonatal and embryo haplotypes. Dark blue represents the homologue inherited from the grandfather of the child and light blue represent the homologue inherited from the grandmother. In Family 7 (**d**), the father inherited the mutant allele from the grandfather and the disease locus crosses the light blue haplotype block indicating the wild-type allele from grandmother was transmitted. **e-f**, Families with parents of the mother for phasing, dark red represent the homologue inherited from the grandfather, and light red represent the homologue inherited from the grandmother. In both families, the mutant allele was inherited from the grandmother.

Supplementary Figure S5. Factors that impact the accuracy cffDNA haplotyping. ~~**Lower informative SNPs inference accuracy near homologous recombination sites.**~~ **a**, Measurement of accuracy for cfDNA haplotyping and illustration of density of SNPs, sequencing depth and FF effects on haplotype accuracy and crossover resolution metrics. **b**, For a 9.5%FF cfDNA samples, homologous recombination accuracy measured in 10Kb bins that flank out from both sides of true (neonatal) crossovers, is shown. On x-axis, 0 denotes the true crossover breakpoint sites. Positive and negative distances correspond to extension to the right and to the left side of breakpoints. **c**, Homologous recombination accuracy measure in terms of haplotype-based informative SNPs. Informative SNPs genotypes inferred from cffDNA haplotyping were evaluated against neonatal genotypes from Family 1_181. x-axis

denotes called informative SNPs in the cfDNA samples and 0 indicates neonatal crossover breakpoint sites. Extending from crossovers, accuracy of informative SNPs to the left (negative values) and to the right side (positive) was measured for maternal and paternal transmissions.

Supplementary Figure S6. Genome-wide haplotyping comparison with reference haplotype. **a**, cffDNA haplotyping results compared to born child haplotype. **b**, cffDNA haplotyping results compared to embryo single-cell haplotype. In all sub figures, paternal inheritance is shown in dark and light blue, and maternal inheritance is shown in dark and light red. Upper tracks are reference (born child/embryo) haplotyping results and lower tracks are cffDNA haplotyping results.

Supplementary Figure S7. Additional informative SNPs improve performance of cffDNA haplarithmisis. Integrating phased type 4 SNPs into maternal informative SNPs show higher accuracy and decreased crossover gaps in cffDNA haplotyping at varying simulated fetal fraction levels.

Supplementary Figure S8. cffDNA haplotyping accuracy of downsampled data. Effect of sequencing depth on the performance of cffDNA haplotyping for a 9.5% FF sample.

Supplementary Figure S9. Patterns of chromosomal anomalies. Assuming 100-fold coverage and 10% fetal fraction, different segmentation schemes are shown in normal,

maternal trisomy, and paternal trisomy in **a** and corresponding FAR values are displayed in **b**.

Supplementary Figure S10. Spike-in samples FAR values deviated from normal values significantly. In the 20% spike-in sample, **a**, For M1 maternal informative SNPs, analysis by unpaired two groups *t*-test of FAR values of chromosome 21 and FAR values of other chromosomes indicating inheritance of maternal H1 showed significant (p-value < 2.22e-16) difference in mean of FAR values. **b**, For M2 maternal informative SNPs, *t*-test also showed significant (p-value < 2.22e-16) difference between mean of chromosome 21 FAR values and other chromosomes FAR values of maternal H1 inheritance. In the 10% spike-in sample, *t*-test showed significant FAR values shifting for M1 maternal informative SNPS shown in **c** (p-value = 1.5e-10) and for M2 maternal informative SNPs shown in **d** (p-value = 2.9e-15), respectively.

Supplementary Figure S11. Simulation of lower fetal fraction effect on paternal haplotyping accuracy. Genome-wide paternal inheritance haplotyping accuracy and crossover resolution of lower fetal fraction.

Supplementary Figure S12. Maternal segmental duplication. a, log2 ratio profile showing ~1.5Mb duplication from maternal gDNA in family 9_074. **b**, log2 ratio profile of maternal copy number presence in cfDNA sample of case 9_074.

Supplementary Figure S13. Reference allele bias in cfDNA samples. Density plot of reference allele frequencies at heterozygous SNPs (both maternal and fetal) for 6 cfDNA

samples. The distribution show centered reference allele frequency to be higher than theoretical value of 0.5.

Supplementary Tables

Supplementary Table S1. cffDNA haplotyping accuracy

Family	Plasma DNA Coverage[1]	Fetal Fraction[2]	No. SNPs		Haplotype accuracy (%) w.r.t. neonatal haplotype		Informative SNPs accuracy (%) w.r.t neonatal SNPs		Haplotype accuracy (%) w.r.t. embryo haplotype		Crossover resolution (Kb)	
			Paternal	Maternal	Paternal	Maternal	Paternal	Maternal	Paternal	Maternal	Paternal	Maternal
1_181	76x	9.5%	34,043	34,362	98.94	94.53	99.70	95.64	99.31	95.69	268.93	818.25
2_186[3]	85x	9.8%	36,579	39,224	-	-	-	-	99.05	93.16	149.09	570.46
3_085	77x	16.5%	38,573	40,887	-	-	-	-	99.42	97.24	131.18	433.74
4_158	64x	8.2%	37,855	-	98.84	-	99.04	-	98.86	-	308.47	-
5_149	96x	6.8%	40,695	-	-	-	-	-	99.77	-	148.52	-
6_150	63x	6.2%	39,237	-	99.00	-	99.40	-	99.15	-	330.53	-
7_054	50x	13.4%	36,558	-	-	-	-	-	99.46	-	162.57	-
8_078	82x	11.4%	-	43,401	-	-	-	-	-	92.68	-	1619.10
9_074	82x	8.4%	-	40,651	-	-	-	-	-	89.41	-	1686.04

[1]Median on-target de-duplicated fragment coverage; [2]Estimated from Type 1 SNPs; [3]Family with consanguinity

Supplementary Table S2. Sequencing or mapping error rate in cfDNA and Spike-in samples

Sample	Error rate (%)
1_181	0.0430
2_186	0.0498
3_085	0.0498
4_158	0.0387
5_149	0.0563
6_150	0.0465
7_054	0.1025
8_078	0.1439
9_074	0.1196
Spike-in_20%	0.0437
Spike-in_10%	0.0445

LICENCE TO PUBLISH

SPRINGER NATURE

Manuscript Number: GIM-D-19-00723

Journal Name: Genetics in Medicine (the "Journal")

Proposed Title of the Article: Non-invasive prenatal diagnosis by genome-wide haplotyping of cell- (the "Article")

Author(s) [Please list all authors, continuing on a separate sheet if necessary]: Huiwen Che, Darine Villela, Eftychia Dimitriadou, Cindy Melotte, Nathalie Brison, Maria Neofytou, Kris Van Den Bogaert, Olga Tsuiko, Koen Devriendt, Eric Legius, Masoud Zamani Esteki, Thierry Voet, Joris Robert (the "Author(s)")

Miscellaneous [for office use only]:

will consider publishing this article, including any supplementary information and graphic elements therein (e.g. illustrations, charts, moving images) (the 'Article'). Headings are for convenience only.

2. Grant of Rights
In consideration of the Licensee evaluating the Article for publication, the Author(s) grant the Licensee the exclusive (except as set out in clauses 3, 4 and 5a) iv) and sub-licensable right, unlimited in time and territory, to copy-edit, reproduce, publish, distribute, transmit, make available and store the Article, including abstracts thereof, in all forms of media of expression now known or developed in the future, including pre- and reprints, translations, photographic reproductions and extensions. Furthermore, to enable additional publishing services, such as promotion of the Article, the Author(s) grant the Licensee the right to use the Article (including the use of any graphic elements on a stand-alone basis) in whole or in part in electronic form, such as for display in databases or data networks (e.g. the Internet), or for print or download to stationary or portable devices. This includes interactive and multimedia use as well as posting the Article in full or in part or its abstract on social media, and the right to alter the Article to the extent necessary for such use. The Licensee may also let third parties share the Article in full or in part or its abstract on social media and may in this context sub-license the Article and its abstract to social media users. Author(s) grant to Licensee the right to re-license Article metadata without restriction (including but not limited to author name, title, abstract, citation, references, keywords and any additional information as determined by Licensee).

3. Self-Archiving
Author(s) are permitted to self-archive a pre-print and an Author's accepted manuscript version of their Article.
a) A pre-print is the Author's version of the Article before peer-review has taken place ("Pre-Print"). Prior to acceptance for publication, Author(s) retain the right to make a Pre-Print of their Article available on any of the following: their own personal, self-maintained website; a legally compliant pre-print server such as but not limited to arXiv and bioRxiv. Once the Article has been published, the Author(s) should update the acknowledgement and provide a link to the definitive version on the publisher's website: "This is a pre-print of an article published in [insert journal title]. The final authenticated version is available online at: https://doi.org/[insert DOI]"
b) An Author's Accepted Manuscript ("AAM") is the version accepted for publication in a journal following peer review but prior to copyediting and typesetting that can be made available under the following conditions:
Author(s) retain the right to make an AAM of their Article available for public release on any of the following 6 months after first publication ("Embargo Period"): their own personal, self-maintained website; their employer's internal website; their institutional and/or funder repositories. AAMs may also be deposited in such repositories immediately on acceptance, provided that they are not made publicly available until after the Embargo Period.
An acknowledgement in the following form should be included, together with a link to the published version on the publisher's website: "This is a post-peer-review, pre-copyedit version of an article published in [insert journal title]. The final authenticated version is available online at: http://dx.doi.org/[insert DOI]"

4. Author's Retained Rights
Author(s) retain the following non-exclusive rights for the published version provided that, when reproducing the Article or extracts from it, the Author(s) acknowledge and reference first publication in the Journal:
a) to reuse graphic elements created by the Author(s) and contained in the Article, in presentations and other works created by them;
b) they and any academic institution where they work at the time may reproduce the Article for the purpose of course teaching (but not for inclusion in course pack material for onward sale by libraries and institutions); and
c) to reproduce, or to allow a third party Licensee to reproduce the Article in whole or in part in any thesis written by the Author(s).
d) to reproduce, or to allow a third party Licensee to reproduce the Article, in whole or in part, in any other type of work (other than thesis) written by the Author(s) for distribution by a publisher after an embargo period of 12 months.

5. Warranties
The Author(s) warrant and represent that:
a) (i) the Author(s) are the sole copyright owners or have been authorised by any additional copyright owner to grant the rights defined in clause 2, (ii) the Article does not infringe any intellectual property rights (including without limitation copyright, database rights or trade mark rights) or other third party rights and no licence from or payments to a third party are required to publish the Article, (iii) the Article has not been previously published or licensed, (iv) if the Article contains materials from other sources (e.g. illustrations, tables, text quotations), Author(s) have obtained written permissions to the extent necessary from the copyright holder(s), to license to the Licensee the same rights as set out in clause 2 but on a non-exclusive basis and without the right to use any graphic elements on a stand-alone basis and have cited any such materials correctly;
b) all of the facts contained in the Article are according to the current body of science true and accurate;
c) nothing in the Article is obscene, defamatory, violates any right of privacy or publicity, infringes any other human, personal or other rights of any person or entity or is otherwise unlawful and that informed consent to publish has been obtained for all research participants;
d) nothing in the Article infringes any duty of confidentiality which any of the Author(s) might owe to anyone else or violates any contract, express or implied, of any of the Author(s). All of the institutions in which work recorded in the Article was created or carried out have authorised and approved such research and publication; and
e) the signatory (the Author or the employer) who has signed this agreement has full right, power and authority to enter into this agreement on behalf of all of the Author(s).

6. Cooperation
a) The Author(s) shall cooperate fully with the Licensee in relation to any legal action that might arise from the publication of the Article and the Author(s) shall give the Licensee access at reasonable times to any relevant accounts, documents and records within the power or control of the Author(s). The Author(s) agree that the distributing entity is intended to have the benefit of and shall have the right to enforce the terms of this agreement.
b) The Author(s) authorise the Licensee to take such steps as it considers necessary at its own expense in the Author(s)' name and on their behalf if the Licensee believes that a third party is infringing or is likely to infringe copyright in the Article including but not limited to initiating legal proceedings.

7. Author List
After signing, changes of authorship or the order of the authors listed will not be accepted unless formally approved in writing by the Licensee.

8. Edits & Corrections
The Author(s) agree(s) that the Licensee may retract the Article or publish a correction or other notice in relation to the Article if the Licensee considers in its reasonable opinion that such actions are appropriate from a legal, editorial or research integrity perspective.

Signed for and on behalf of the Author(s):

Print Name: PROF. JORIS VERMEESCH

Date: 19/11/2019

(PLEASE NOTE, ONLY HANDWRITTEN SIGNATURES ARE ACCEPTED)

Address:
HERESTRAAT 49, O2 GASTHUISBERG
3000 LEUVEN, BELGIUM

Corresponding Author Name: Joris Robert Vermeesch
Manuscript Number: GIM-D-19-00723

Reporting Checklist* (Please see page 3 for instructions on uploading this file)

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. **Please respond completely to all questions relevant to your manuscript. For sections that are not applicable please fill in NA.** For more information, please read the journal's Guide to Authors.

⬓ Check here to confirm that the following information is available in the Material & Methods section:

- the **exact sample size (*n*)** for each experimental group/condition, given as a number, not a range;
- a **description of the sample collection** allowing the reader to understand whether the samples represent **technical or biological replicates** (including how many animals, litters, culture, etc.);
- a **statement of how many times the experiment shown was replicated in the laboratory**;
- **definitions of statistical methods and measures**: (For small sample sizes (n<5) descriptive statistics are not appropriate, instead plot individual data points)
    - very common tests, such as *t*-test, simple χ^2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
    - are tests one-sided or two-sided?
    - are there adjustments for multiple comparisons?
    - **statistical test results**, e.g., ***P* values**;
    - definition of **'center values'** as **median or mean**;
    - definition of **error bars as s.d. or s.e.m. or c.i.**

Please ensure that the answers to the following questions are reported both **in the manuscript itself and in the space below.** We encourage you to include a specific subsection in the methods section each for statistics, reagents and animal models. Below, provide the text as it appears in the manuscript as well as the page number.

Statistics and general methods	**Text AND page number from manuscript**
1. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size? (Give text and page #)	Not available
For animal studies, include a statement about sample size estimate even if no statistical methods were used.	Not available
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established? (Give text and page #)	Not available
3. If a method of randomization was used to determine how samples/animals were allocated to experimental groups and processed, describe it. (Give text and page #)	Not available
For animal studies, include a statement about randomization even if no randomization was used.	Not available

4. If the investigator was blinded to the group allocation during the experiment and/or when assessing the outcome, state the extent of blinding. (Give text and page #)	Not available
For animal studies, include a statement about blinding even if no blinding was done.	Not available
5. For every figure, are statistical tests justified as appropriate?	Yes
Do the data meet the assumptions of the tests (e.g., normal distribution)?	Not available
Is there an estimate of variation within each group of data?	Not available
Is the variance similar between the groups that are being statistically compared? (Give text and page #)	Not available

Reagents	**Text AND page number from manuscript**
6. Report the source of antibodies (vendor and catalog number)	Not available
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination	Not available

Animal Models	**Text AND page number from manuscript**
8. Report species, strain, sex and age of animals	Not available
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	Not available

10. We recommend consulting the ARRIVE guidelines (*PLoS Biol.* **8**(6), e1000412,2010) to ensure that other relevant aspects of animal studies are adequately reported.

Human subjects	**Text AND page number from manuscript**
11. Identify the committee(s) approving the study protocol.	Page6. This study was approved by the local Ethical Committees of the University Hospital Leuven (S59324).
12. Include a statement confirming that informed consent was obtained from all subjects.	Page6. Women with a successful pregnancy following PGT-M were recruited at the UZ Leuven Hospital, with informed
13. For publication of patient photos, include a statement confirming that consent to publish was obtained. For more information, please see http://www.icmje.org/recommendations/browse/roles-and-responsibilities/protection-of-research-participants.html.	Not available
14. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent).	Not available

15. For phase II and III randomized controlled trials, please refer to the CONSORT statement and submit the CONSORT checklist with your submission.

16. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines.

Data deposition	**Text AND page number from manuscript**
17. Provide accession codes for deposited data. Data deposition in a public repository is recommended for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Microarray data	Page21. EGAS00001003634

Deposition is strongly recommended for many other datasets for which structured public repositories exist; more details on our data policy are available in the Guide to Authors. We encourage the provision of other source data in supplementary information or in unstructured repositories such as Figshare and Dryad. We encourage publication of Data Descriptors (see Scientific Data) to maximize data reuse.

18. If computer code was used to generate results that are central to the paper's conclusions, include a statement in the Methods section under "**Code availability"** to indicate whether and how the code can be accessed. Include version information as necessary and any restrictions on availability.	The code is deposited in GITHUB. https://github.com/vermeeschlab/cfDNAhaplotyping The code will become available under a license for research.

* If an error is shown on the PDF built in Editorial Manager after including this Reporting Checklist take the following steps:
1. Open the original filled GIM-reporting-checklist.pdf with Mac Preview App
2. Goto Tray -> Export as PDF
3. Save as: at your desk (local computer) with a different file name
4. Close the app
5. In Editorial Manager: Edit the Submission which displayed Error or Incomplete Status and remove the file, then upload the new one, and again select to build the merged PDF

COLOR ARTWORK FORM

This form must be completed for all papers. We will be unable to process your paper through to production until we receive instructions concerning color files. Please upload this form with your revised files.

JOURNAL: ***Genetics in Medicine***

ARTICLE TITLE Non-invasive prenatal diagnosis by genome-wide haplotyping of cell-free plasma DNA

MANUSCRIPT NUMBER GIM-D-19-00723

CORRESPONDING AUTHOR NAME Joris Robert Vermeesch

Genetics in Medicine has charges for figures printed in color: $500 for the first color figure and $250 for each subsequent figure. Current ACMG members (excluding Student Members) who are first or senior/corresponding authors are exempt,* as are authors who have opted for Open Access.*

Please note that figures can appear online in color in the HTML version of your manuscript, and in black and white in the PDF/print version of the manuscript. Color figures will also be at the discretion of the editorial office.

Please check:

☐ Yes, my manuscript contains material that must be printed in color. I agree to pay the color charges in full and hereby authorize Nature Publishing Group to invoice me for the cost of reproducing color artwork in print.

☐ Yes, my manuscript contains material that should be in color in the online HTML version, but in black and white in the PDF/print version of the manuscript. No charges incurred.

☒ Yes, my manuscript contains material that must be printed in color, but I have chosen Open Access for my article or am an ACMG member, so am exempt.

☐ No, my manuscript does not contain material that must be printed in color.

Which figures should be printed in color? (e.g. Figures 1a, 2, 3b)

College Member's Name: KU Leuven

Membership will be verified by the ACMG, and false claims will be liable for the full color charge rate. Signed completion of this form constitutes a full and total acceptance of the terms listed. College membership has no bearing on the review process at *Genetics in Medicine*, and papers from members and nonmembers alike will be given equal consideration.

Color figures will be set close to the citation and in the best possible position.

20% VAT will be added to the total charge amount upon invoicing. This applies to all EU authors who do not provide a valid VAT number upon returning this form. Customers outside of the EU will not be charged VAT but local taxes will be added where applicable.

Signature: _____ VAT no. (if applicable): BE 0419.052.173

Print Name: Prof. JORIS VERMEESCH Date: 19/11/2019

Updated 8/16/16

nature publishing group npg