Clinical Chemistry 63:2
513–524 (2017)

Universal Haplotype-Based Noninvasive Prenatal Testing for Single Gene Diseases

Winnie W.I. Hui,[1] Peiyong Jiang,[1,2] Yu K. Tong,[1] Wing-Shan Lee,[1] Yvonne K.Y. Cheng,[3] Maria I. New,[4] Rezan A. Kadir,[5,6] K.C. Allen Chan,[1,2] Tak Y. Leung,[3] Y.M. Dennis Lo,[1,2] and Rossa W.K. Chiu[1,2*]

BACKGROUND: Researchers have developed approaches for the noninvasive prenatal testing of single gene diseases. One approach that allows for the noninvasive assessment of both maternally and paternally inherited mutations involves the analysis of single nucleotide polymorphisms (SNPs) in maternal plasma DNA with reference to parental haplotype information. In the past, parental haplotypes were resolved by complex experimental methods or inferential approaches, such as through the analysis of DNA from other affected family members. Recently, microfluidics-based linked-read sequencing technology has become available and allows the direct haplotype phasing of the whole genome rapidly. We explored the feasibility of applying this direct haplotyping technology in noninvasive prenatal testing.

METHODS: We first resolved the haplotypes of parental genomes with the use of linked-read sequencing technology. Then, we identified SNPs within and flanking the genes of interest in maternal plasma DNA by targeted sequencing. Finally, we applied relative haplotype dosage analysis to deduce the mutation inheritance status of the fetus.

RESULTS: Haplotype phasing and relative haplotype dosage analysis of 12 out of 13 families were successfully achieved. The mutational status of these 12 fetuses was correctly classified.

CONCLUSIONS: High-throughput linked-read sequencing followed by maternal plasma-based relative haplotype dosage analysis represents a streamlined approach for noninvasive prenatal testing of inherited single gene diseases. The approach bypasses the need for mutation-specific assays and is not dependent on the availability of DNA from other affected family members. Thus, the approach is universally applicable to pregnancies at risk for the inheritance of a single gene disease.

The presence of cell-free fetal DNA in maternal plasma *(1)* offers a noninvasive approach for prenatal diagnosis. Maternal plasma DNA analysis for the screening of common fetal chromosomal aneuploidies has been achieved with high degree of accuracy *(2, 3)* resulting in substantial reductions in the number of invasive prenatal diagnostic procedures performed.

Apart from fetal aneuploidies, single gene disease is the other reason why some pregnant women consider prenatal diagnosis. Since fetal DNA is present in a background of maternal DNA *(4)*, early work for the noninvasive determination of single gene disease inheritance focused on the analysis of paternally transmitted fetal-specific sequences or mutations that could be distinguished from the maternal genome. For example, the detection of chromosome Y sequences in maternal plasma allowed accurate fetal sex determination and hence served as a means to evaluate the risk of a fetus for having a sex-linked disorder *(5–7)*. The presence or absence of paternally inherited mutant alleles in maternal plasma has been applied to the noninvasive assessment of paternally inherited autosomal dominant diseases or for the exclusion of the fetus being affected by an autosomal recessive disease *(8–10)*.

To assess the fetal inheritance of maternally transmitted mutations, approaches have been developed to compare the relative amounts of the mutant and wild-type alleles or haplotypes in maternal plasma. The relative mutation dosage approach directly measures the number of DNA molecules in maternal plasma that carry the mutant or wild-type alleles. For a mother who is a carrier of a mutation, equal amounts or skewed amounts

[1] Department of Chemical Pathology, The Chinese University of Hong Kong, Hong Kong SAR, China; [2] Centre for Research into Circulating Fetal Nucleic Acids, Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China; [3] Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong SAR, China; [4] Department of Pediatrics, Icahn School of Medicine at Mount Sinai, NY; [5] Department of Obstetrics and Gynaecology, Royal Free London NHS Foundation Trust, London, UK; [6] Katharine Dormandy Haemophilia Centre and Thrombosis Unit, Royal Free London NHS Foundation Trust, London, UK.

* Address correspondence to this author at: Department of Chemical Pathology, Prince of Wales Hospital, 30-32 Ngan Shing St., New Territories, Hong Kong SAR, China. Fax +852-2636-5090; e-mail rossachiu@cuhk.edu.hk.
Received October 18, 2016; accepted November 17, 2016.
Previously published online at DOI: 10.1373/clinchem.2016.268375

between the 2 alleles in maternal plasma would provide an indication of whether the fetus is heterozygous or homozygous for either allele, respectively *(11, 12)*.

The relative haplotype dosage (RHDO)[7] approach, on the other hand, allows the deduction of the fetal genotype by measuring the relative counts of single nucleotide polymorphism (SNP) alleles on haplotypes linked with the mutant allele and wild-type allele in maternal plasma DNA *(13)*. This method allows the indirect measurement of mutations that are more challenging to be detected by direct mutation-specific assays, such as gene deletion, inversion, mutations in repetitive elements and homologous genes *(14)*. The RHDO method could be applied in a genome-wide *(13)* or a targeted fashion specifying the analysis for particular loci *(14, 15)*.

In RHDO analysis, maternal haplotype information is required. However, haplotype phasing strategies used in previous studies were complicated and laborious. Methods to determine haplotype information include inferential statistical analysis and direct experimental techniques. By genotyping genomic DNA of trios, including the father, mother and an affected proband in the family, SNPs linked with mutation sites could be identified and thus haplotypes could be deduced *(14)*. This approach restricts the application of the testing to families with a previously affected family member whose DNA is available. Alternatively, haplotypes could be preconstructed by population-based inference *(16)* or reconstructed from genomic DNA of an individual by methods such as clone pool dilution sequencing *(17)*, contiguity-preserving transposition sequencing *(18)* and HaploSeq *(19)*. However, these techniques require intricate experimental protocols or reagents that are not yet widely commercially available *(20)*.

Recently, a direct haplotype phasing approach using microfluidics-based linked-read sequencing technology became available *(21)*. Long input DNA molecules are partitioned into droplets and transformed into short barcoded fragments for sequencing. Identical barcodes are used to identify short fragments that originate from the same long genomic DNA molecule. Computational algorithm is then used to reconstruct the short-read sequencing data into long-range haplotype information.

In this study, we applied linked-read sequencing technology to directly generate haplotype-resolved genome sequence from parental DNA. Maternal plasma DNA sequencing data were interpreted with the parental haplotype information to deduce the mutational status of the fetus. We assessed the feasibility of using this protocol for the noninvasive prenatal assessment of a number of autosomal and X-linked diseases, showing that this streamlined approach enabled noninvasive detection of single gene disease inheritance without the need to design bespoke assays to assess mutations on a case-by-case basis *(22, 23)* and only required the use of specimens from the parents.

[7] Nonstandard abbreviations: RHDO, relative haplotype dosage; SNP, single nucleotide polymorphism; KS, Kolmogorov–Smirnov; SPRT, sequential probability ratio test; CAH, congenital adrenal hyperplasia; EVC, Ellis-van Creveld syndrome; chr, chromosome.

Materials and Methods

GENERAL PRINCIPLES

Parental haplotypes were first determined using microfluidics-based linked-read sequencing *(21)* on blood cell DNA obtained from the pregnant woman and her male partner. Maternal plasma DNA were then subjected to targeted sequencing and SNP alleles located upstream and downstream of a disease locus were identified. The haplotype origin of each SNP allele is deduced. A statistical comparison between the abundance of plasma DNA molecules derived from the 2 maternal haplotypes was performed. Another statistical comparison between the abundance of plasma DNA molecules derived from the 2 paternal haplotypes was performed. The fetal genotype was then deduced based on the 2 sets of statistical results. Design of the target capture probes for targeted sequencing is provided in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol63/issue2.

SAMPLE COLLECTION

Patients were recruited at the Prince of Wales Hospital, Hong Kong; Royal Free Hospital, London, UK; and Mount Sinai School of Medicine, New York, NY, with informed consent. Ethics approvals from all respective institutional boards were granted. 5–10 mL maternal blood samples were collected before any invasive procedures during pregnancy. Paternal and maternal blood samples were centrifuged at 1600*g* for 10 min at 4 °C, and the plasma portion was recentrifuged at 16 000*g* for 10 min at 4 °C *(24)*. Plasma, buffy coat and genomic DNA were transferred to the Hong Kong laboratory at the Department of Chemical Pathology at The Chinese University of Hong Kong for analysis. The paternal and maternal buffy coat DNA processing and the plasma DNA processing are described in the online Supplemental Methods file.

SEQUENCE READ ALIGNMENT AND MEASUREMENT OF FRACTIONAL FETAL DNA CONCENTRATION

Information describing the sequence read alignment and the measurement of fractional fetal DNA concentration is provided in the online Supplemental Methods file.

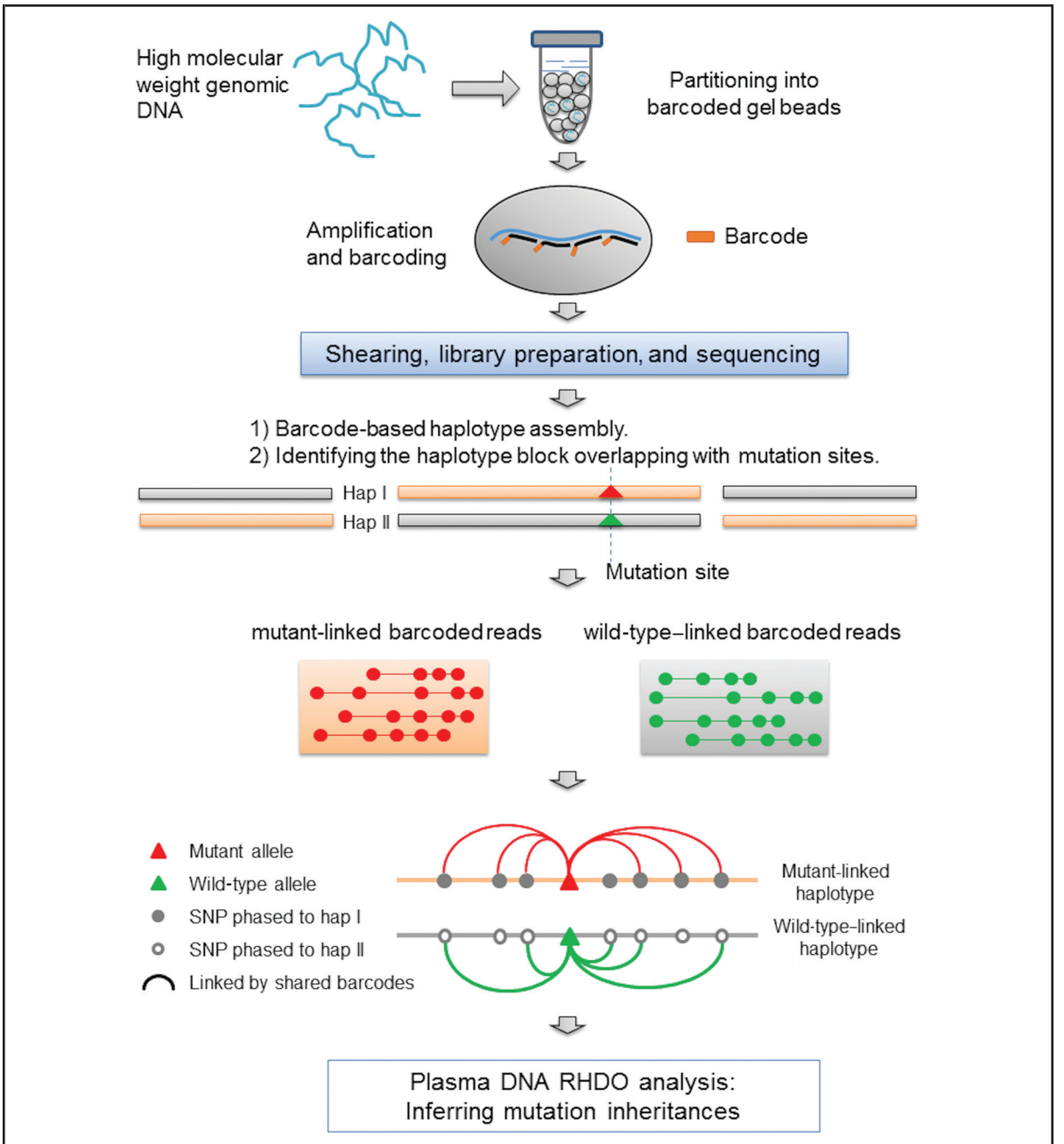

Fig. 1. Schematic diagram of the principle of haplotype phasing.

Long DNA molecules were partitioned into gel beads and amplified by unique 10x™ barcoded primers. Barcode-tagged short reads were sequenced. Sequenced reads that shared the same barcode with the reads carrying mutant allele were linked (mutant-linked barcode reads) and phased to the same haplotype. Wild-type–linked barcode reads were phased to the opposite haplotype. SNP information on the mutant- or wild-type–linked haplotype were extracted for RHDO analysis.

HAPLOTYPE PHASING

Haplotype phasing of genomic DNA was achieved by linking short read sequencing data to provide long range genetic information (Fig. 1). From the buffy coat sequencing data, barcode information of each sequenced read was used to link the short sequenced reads into original long input molecules. With sufficient dilution, the chance of having 2 distinct long DNA molecules that

cover a genomic locus with opposing haplotype in the same gel bead is very low. Therefore, reads that shared the same barcode with the ones carrying mutant alleles were phased to the same haplotype (termed *Hap*I or mutant-linked haplotype) and those which shared the same barcode with the ones carrying wild-type alleles were phased to the opposite haplotype (termed *Hap*II or wild-type–linked haplotype). SNPs linked on the same haplotypes with the mutant and wild-type alleles were noted and were used for the subsequent maternal plasma DNA RHDO analysis.

KOLMOGOROV-SMIRNOV TEST OF PATERNALLY TRANSMITTED AUTOSOMAL MUTATIONS

Informative SNPs where the mother was homozygous and the father was heterozygous were analyzed. If the fetus had inherited the mutation from father, the paternal-specific SNP alleles that could be detected in maternal plasma would belong to the paternal mutant-linked haplotype as identified by the haplotype analysis of the paternal DNA. Kolmogorov–Smirnov (KS) test was applied to determine whether there was a statistical difference of allelic counts between the 2 paternal haplotypes. Read counts of paternal-specific alleles between paternal haplotypes were respectively accumulated until a mutant-linked haplotype or a wild-type–linked haplotype was classified *(14)*. To minimize stochastic influences, the haplotype block was required to fit certain criteria: the number of SNPs in test region $\geq$25; the cumulative difference between 2 haplotypes $>$0.53%; and the *P* value of the KS test $<$0.05 (14). With the use of the KS test, we could minimize the analytical errors caused by sequencing errors by providing an unbiased statistical comparison between the 2 haplotypes. As a result, locations of recombination events that may occur between paternal haplotypes could be pinpointed with higher precision.

RHDO ANALYSIS OF MATERNALLY TRANSMITTED AUTOSOMAL MUTATIONS

RHDO analysis based on sequential probability ratio test (SPRT) classification was performed to deduce the fetal inheritance of the maternally transmitted mutations *(13, 14)*. Informative SNPs where mother was heterozygous and father was homozygous were analyzed. Each SNP was classified as type α or type β. For type α SNPs, the paternal alleles were identical to the maternal alleles on the maternal mutant-linked haplotype. If the fetus had inherited the mutant allele, an overrepresentation of mutant-linked haplotype would be observed in maternal plasma DNA. In contrast, if the fetus had inherited the wild-type allele, there would be no overrepresentation of either one of the maternal haplotypes. For type β SNPs, the paternal alleles were identical to the maternal alleles on the maternal wild-type–linked haplotype, i.e., haplotype linked with the wild-type allele. If the fetus had inherited the wild-type haplotype, an overrepresentation of wild-type–linked haplotype would be observed. On the other hand, if the fetus had inherited the mutant allele, both haplotypes would be equally represented. RHDO analysis involved a statistical evaluation of dosage balance or imbalance between alleles to determine the haplotype block inherited *(13)*.

Further information about the RHDO analysis of maternally transmitted autosomal mutations can be found in the online Supplemental Methods file.

STATISTICAL ANALYSIS FOR THE ASSESSMENT OF X-LINKED INHERITANCE

Informative SNPs on chromosome X where mother was heterozygous were analyzed. If the fetus had inherited the mutation, there would be an overrepresentation of mutant-linked haplotype in maternal plasma DNA analysis. If the fetus had inherited the wild-type allele, there would be an underrepresentation of mutant-linked haplotype. SPRT was used to test the 2 alternative hypotheses: (*a*) the mutant allele was overrepresented when compared to the wild-type allele, and (*b*) the mutant allele was underrepresented when compared to the wild-type allele *(12)*.

Results

STUDY PARTICIPANTS

Thirteen families at risk for a fetus with congenital adrenal hyperplasia (CAH), β-thalassemia, Ellis-van Creveld syndrome (EVC), hemophilia or Hunter syndrome were recruited. Except for the pregnancy affected by EVC, each of the recruited families had a known family history of the disease for which conventional prenatal diagnosis was sought. For the EVC case, ultrasound examination revealed multiple structural abnormalities that led to the suspicion of EVC. The disease status of the fetus was determined by conventional prenatal assessment based on mutational analysis of the parental DNA and the fetal DNA obtained by chorionic villus sampling or amniocentesis or after delivery by cord blood or newborn DNA analysis. The mutational status of the studied cases is listed in Table 1.

SEQUENCING

For the CAH families, linked short reads were prepared from the parental buffy coat DNA that were target captured and sequenced to an average of 646-fold haploid human coverage. For the other families, genome-wide sequencing of the linked short reads prepared from the parental buffy coat DNA was performed to a mean of 34-fold haploid coverage. N50 phase block length of the parental DNA samples

Table 1. Clinical information of the recruited families.

Family	Diseases	Gene	Location of gene	Genotypes			Fetal sex	Gestation age (weeks)[a]	Fetal DNA concentration (%)
				Mother	Father	Fetus			
A	CAH	*CYP21A2*	chr6	del/nL[b]	Int2/nL	nL/nL	F	8 1/7	5.4
B	CAH	*CYP21A2*	chr6	Int2/nL	Ex3/nL	Ex3/Int2	M	11 1/7	8.8
C	CAH	*CYP21A2*	chr6	del/nL	del/nL	nL/nL	F	9 6/7	8.5
D	CAH	*CYP21A2*	chr6	del/nL	L307frameshift/R356W	L307fs/del	F	8 2/7	15.3
E	Beta thalassemia	*HBB*	chr11	c.316-197C>T/nL	c.126_129delCTTT/nL	c.316-197C>T/nL	F	17 3/7	4.7
F	EVC	*EVC* and *EVC2*	chr4	c.1006-12A>G/nL in *EVC2* and c.83T>C/nL in *EVC*	c.871-2_894del/nL in *EVC2*	c.1006-12A>G/c.871-2_894del in *EVC2* and c.83T>C/nL in *EVC*	F	20 3/7	14.7
G	Hemophilia A	*F8*	chrX	c.1898T>G/nL	N/A	nL	M	21 4/7	23.1
H	Hemophilia A	*F8*	chrX	c.71A>G/nL	N/A	nL	M	34	11.8
I	Hemophilia A	*F8*	chrX	c.1292T>C/nL	N/A	c.1292T>C	M	31	12.3
J	Hemophilia B	*F9*	chrX	c.802T>A/nL	N/A	c.802T>A	M	21 2/7	22.2
K	Hemophilia B	*F9*	chrX	c.509G>A/nL	N/A	nL	M	30 6/7	10.1
L	Hemophilia B	*F9*	chrX	c.391 + 10T>G/nL	N/A	c.391 + 10T>G	M	2nd trimester	21.1
M	Hunter syndrome	*IDS* and *IDS2*	chrX	Heterozygous *IDS/IDS2* rearrangement	N/A	Hemizygous *IDS/IDS2* rearrangement	M	13 1/7	12.6

[a] At the time of blood sampling for analysis.
[b] del, 30-kb large gene deletion; Int2, intron 2, c.293-13A/C>G; ex3, exon3, c.332_339del; nL, normal allele; N/A, not applicable.

Table 2. Haplotype phasing data.

Family	Sample	Phase block across target region	Length of phase block across target region (bases)	No. of SNPs across target region
A	Mother	chr6:27979631-32679591	4699960	4519
	Father	chr6:24958611-32429077	7470466	4631
B	Mother	chr6:24958611-32414273	7455662	4079
	Father	chr6:24985475-32414273	7428798	3595
C	Mother	chr6:24957224-38920455	13963231	9106
	Father	chr6:29293472-32431785	3138313	2774
D	Mother	chr6:24958611-32412539	7453928	4152
	Father	chr6:24957224-32456672	7499448	3891
E	Mother	chr11:3628799-12689249	9060450	9468
	Father	chr11:1868168-22588088	20719920	17163
F	Mother	chr4:3570130-8737592	5167462	5292
	Father	chr4:3671398-7773524	4102126	4274
G	Mother	chrX:153020246-154404181	1383935	448
H	Mother	chrX:154210627-155173951	963324	23
I	Mother	chrX:147798372-155025884	7227512	3188
J	Mother	chrX:135540724-139774089	4233365	1163
K	Mother	chrX:134856212-153728057	18871845	8767
L	Mother	chrX:130630587-146943274	16312687	7354
M	Mother	chrX:147662589-152231103	4568514	1235

ranged from 3–14 Mb with >94% of SNPs phased. N50 is an indicator of haplotyping performance and defined as the block length at which the sum of block length of that block and larger blocks represents 50% of the overall phased sequence *(20, 21)*. The mean sequencing depth of maternal plasma DNA was 275-fold. Key summary statistics of the sequencing data are shown in online Supplemental Tables 1 and 2.

PRENATAL ASSESSMENT FOR AUTOSOMAL RECESSIVE DISEASES

Families A to F each presented for prenatal assessment of an autosomal recessive disease. The mutant-linked and the wild-type–linked haplotypes for the mother as well as the father were successfully determined for each case (Table 2 and online Supplemental Table 1). The fetal inheritance of the maternal and paternal haplotypes was determined through statistical comparisons between the maternal plasma DNA sequencing reads. Results for each case are shown in Fig. 2. The deduced fetal genotypes were concordant with the results of the conventional diagnostic tests.

As an illustration, the father in family A was a carrier of a point mutation while the mother was a carrier of a 30-kb deletion at the cytochrome P450 family 21 subfamily A member 2 (*CYP21A2*)[8] locus (Table 1). A maternal blood sample was collected at the gestational age of 8 weeks and 1 day. We first resolved the haplotype of the parents from linked-read sequencing data of the parental buffy coat DNA. For maternal haplotyping, we phased the SNP alleles detected from sequenced reads that showed the same barcode as the wild-type allele without the 30-kb deletion as haplotype II. Then, we inferred the other alleles as derived from the haplotype linked with the 30-kb deletion (Fig. 3A). The phased maternal haplotype block across the target gene was around 4.7 Mb in length and contained 4519 informative SNPs for subsequent maternal plasma RHDO analysis (Table 2). The paternal point mutation was located on chromosome (chr) 6, at genomic coordinate 32006858 (GRCh37/hg19). Sequenced reads that shared the same barcodes with the ones containing the paternal mutant allele were phased to one haplotype (*Hap*III). Those shared the same barcodes with reads carrying wild-type alleles were

[8] Human genes: *CYP21A2*, cytochrome P450 family 21 subfamily A member 2; *IDS*, iduronate 2-sulfatase; *IDSP1* or *IDS2*, iduronate 2-sulfatase pseudogene 1; *CYP21A1P*, cytochrome P450 family 21 subfamily A member 1, pseudogene; *F8*, coagulation factor VIII; *EVC2*, EvC ciliary complex subunit 2; *EVC*, EvC ciliary complex subunit 1; *HBB*, hemoglobin subunit beta; *F9*, coagulation factor IX.

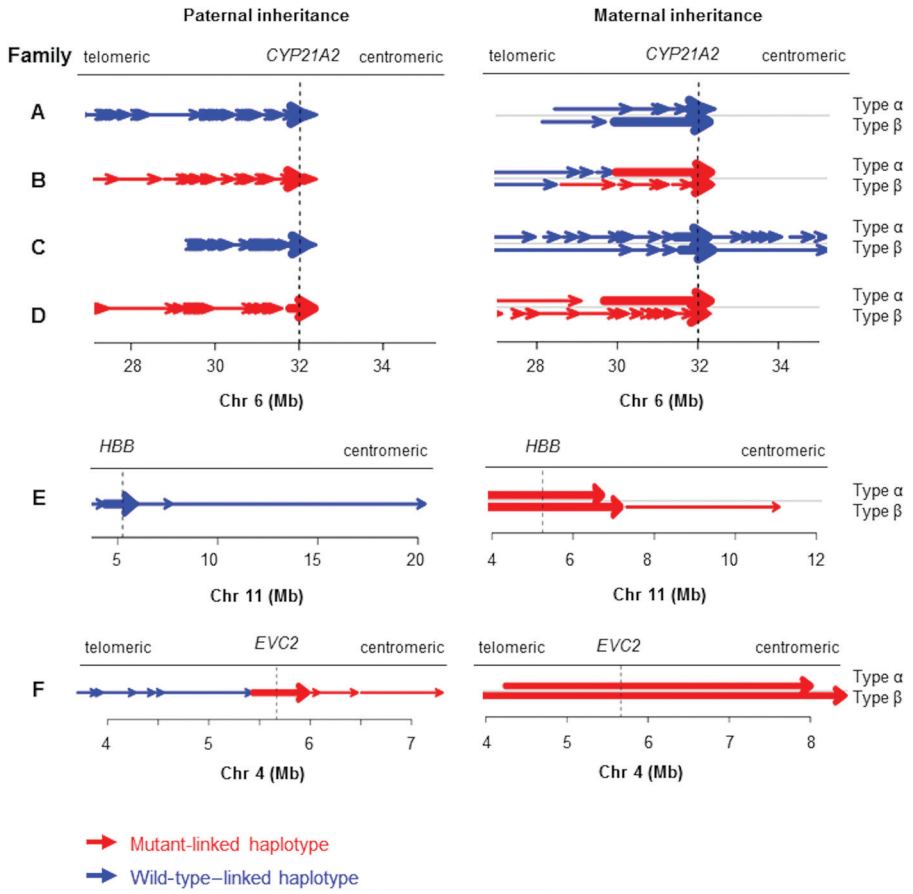

Fig. 2. Fetal haplotype analyses in families A to F.

The analysis started from the SNPs flanking the mutation site and then extended towards the telomeric and centromeric directions. Haplotype block is denoted by an arrow. The tail and tip of the arrow indicate the start and end positions of a haplotype block. The fetal inheritance of which maternal haplotype was determined by RHDO analysis. The fetal inheritance of which paternal haplotype was determined by KS test analysis. A red arrow indicates the inheritance of the mutant-linked haplotype and a blue arrow indicates the inheritance of wild-type–linked haplotype. RHDO analysis was only performed for the EvC ciliary complex subunit 2 (*EVC2*) locus for Family F. EVC syndrome is an autosomal recessive disease caused by mutations in the EvC ciliary complex subunit 1 (*EVC*) or *EVC2* genes and both parents were carriers for mutations on *EVC2*.

phased to the opposite haplotype (*Hap*IV) (Fig. 3B). The phased haplotype block across the target gene was around 7.5 Mb in length and contained 4631 informative SNPs for subsequent maternal plasma KS test analysis (Table 2).

To determine the fetal inheritance of the maternal mutation, we counted the number of plasma DNA molecules carrying informative SNP alleles. Then, we evaluated the haplotype dosage balance or imbalance of type α and type β SNPs with SPRT classification and deduced the haplotype block inherited by the fetus as stated in the Materials and Methods section. A total of 108 type α SNPs and 92 type β SNPs were identified and they were counted separately in the SPRT classification (see online Supplemental Table 3). For type α SNPs, an equal representation of both haplotypes was observed in 6 SPRT classifications. For type β SNPs, an overrepresentation of wild-type–linked haplotype was observed in 2 SPRT classifications. Both analyses indicated that the fetus had inherited the wild-type–linked haplotype from the mother.

To determine the fetal inheritance of the paternal mutation, 2863 informative SNPs within the targeted *CYP21A2* region were detected in maternal plasma. 65 KS tests were done across the locus (see online Supplemental Table 3). Each KS test reached statistical significance (interquartile range of the *P* values $= 10^{-10}$–10^{-4}; minimal cumulative difference between 2 haplotypes $>0.53\%$) indicating that there were more paternal-specific alleles on the wild-type–linked haplotype than those on the mutant-linked haplotype in maternal plasma. The KS test analysis supported the conclusion

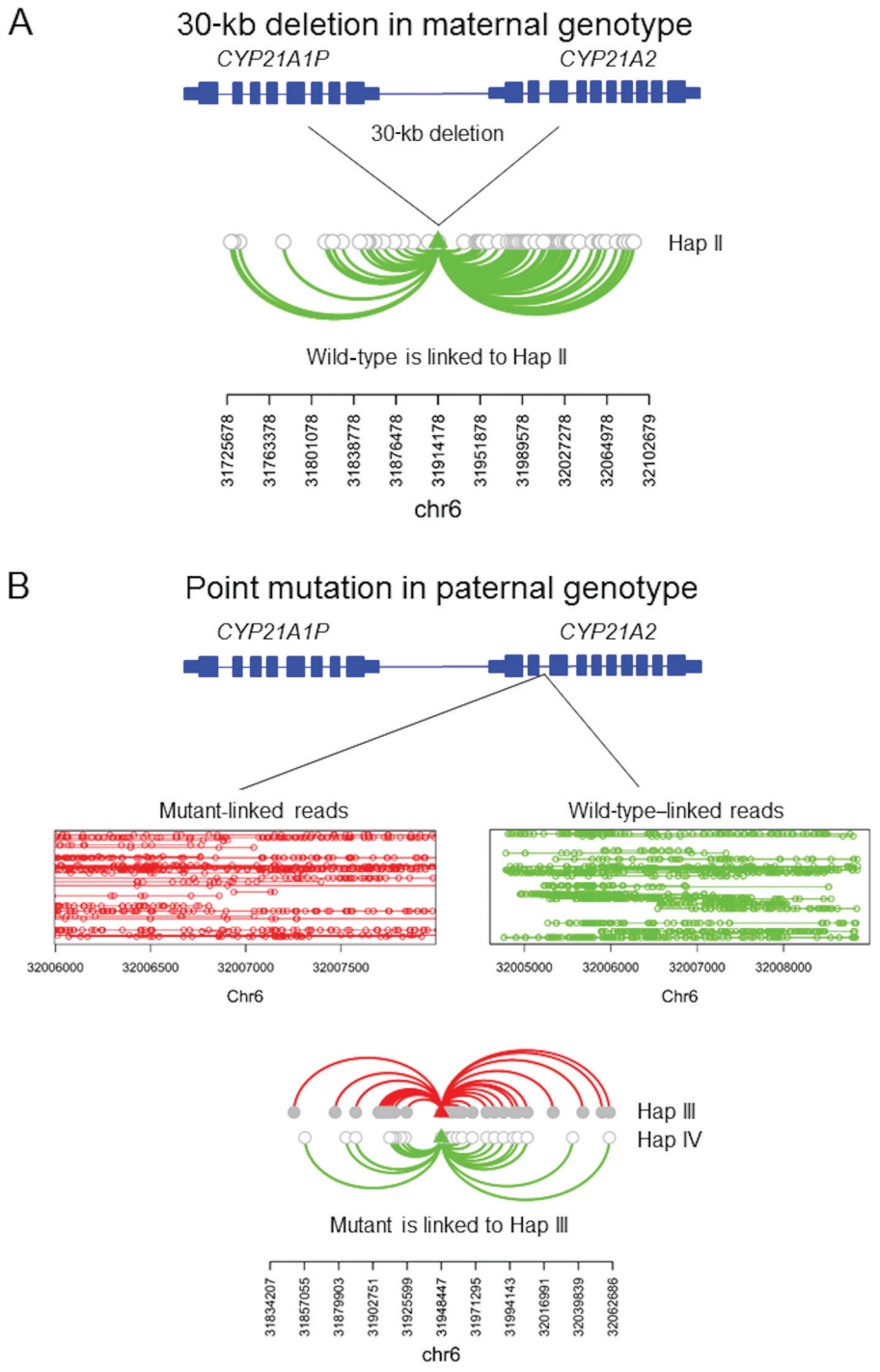

Fig. 3. Haplotype linkage to mutation site in family A.

(A) The mother was a carrier of a 30 kb deletion. SNPs within sequenced reads that shared the same barcode as reads without the 30 kb deletion was considered to be linked to *Hap*II, the wild-type–linked haplotype. Reads containing the alternative alleles were assigned to *Hap*I, mutant-linked haplotype. (B) The father was a carrier of a point mutation. SNPs found on the mutant-linked reads were phased to one haplotype (*Hap*III) and those on the wild-type–linked reads were phased to another (*Hap*IV).

that the fetus had inherited the wild-type–linked haplotype from the father. We therefore concluded that the fetus did not inherit any of the parental mutations and was not affected by CAH.

The same processes were applied to families B to F and the deduced fetal genotypes and hence the disease status were concordant with the conventional prenatal diagnostic results. It is particularly noteworthy that a change in RHDO inheritance was observed in the plasma DNA data for family B and F (Fig. 2). In family B, the maternal haplotype inherited by the fetus deduced from the RHDO analysis changed from wild-type–linked to mutant-linked at around 28–30 Mb on chromosome 6. In family F, there is a shift in the deduced paternal haplotype inherited by the fetus from wild-type–linked to mutant-linked at around 5–5.5 Mb on chromosome 4. In the Figs., a change in the color of the arrows between blue and red indicates the location where a recombination is suspected. The suspected recombinations were confirmed by sequencing the chorionic villus and amniotic fluid samples, respectively.

PRENATAL ASSESSMENT OF X-LINKED DISEASES

Families G to L had a family history of hemophilia A or B. Family M had a family history for Hunter syndrome. Since males are hemizygous for chromosome X, only maternal haplotype analysis and fetal inheritance of the maternal X-linked mutations were performed (Fig. 4). In family G, the mother was a carrier of a point mutation on coagulation factor VIII (*F8*). Haplotypes were constructed from heterozygous SNPs on chromosome X detected from maternal genomic DNA and linkage to the mutant or wild-type allele was determined. The length of the reconstructed haplotype block was 1.4 Mb and contained 448 informative SNPs for inheritance analysis. Because the maternal DNA was subjected to genome-wide sequencing while targeted sequencing was performed for the maternal plasma sample, only 6 of the informative SNPs were detected within the target region in maternal plasma. Nonetheless, 1 SPRT classification spanning the mutation site was achieved. The result showed an underrepresentation of informative SNPs linked with the mutant allele and indicated that the fetus had inherited the wild-type allele from the mother.

In family H, the maternal haplotype was successfully resolved. However, this particular mutation was in an SNP-depleted repeat region and our capture probes were not specifically designed to target regions spanning this mutation site. Also, the maternal plasma volume for DNA extraction was only 0.75 mL, which was much lower than an average of 3.68 mL plasma for the other samples, and this may reduce the DNA amount for RHDO analysis. There were therefore not enough informative SNP data for RHDO classification.

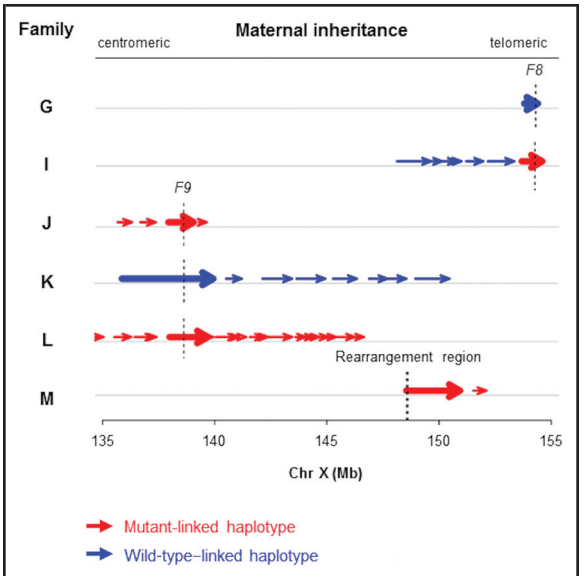

Fig. 4. Fetal haplotype analyses in families G to M.

Since males were hemizygous for chromosome X, we only analyzed maternal inheritance of the X-linked mutations. The analysis started from the SNPs flanking the mutation site and then extended towards the telomeric and centromeric directions. The haplotype block inherited is denoted by an arrow. The tail and tip of the arrow indicate the start and end position of a haplotype block. The fetal inheritance of which maternal haplotype was classified by RHDO analysis. A red arrow infers that an overrepresentation of mutant-linked SNP alleles in the maternal plasma DNA was classified and indicates that the fetus had inherited the mutant-linked haplotype. A blue arrow infers that an overrepresentation of wild-type–linked SNP alleles in the maternal plasma DNA was classified and indicates that the fetus had inherited the wild-type–linked haplotype.

A recombination event was suspected from the maternal plasma DNA analysis performed for family I. The recombination was subsequently confirmed by targeted sequencing of placental DNA. Maternal haplotype analysis and maternal plasma RHDO assessment were successfully performed for families I to L. The deduced fetal genotypes were concordant with the conventional diagnostic results.

In family M, the mother was heterozygous for an iduronate 2-sulfatase (*IDS*/*IDS2*) gene rearrangement. PCR amplification and restriction fragment length polymorphism analysis of maternal DNA and chorionic villi DNA identified a recombination that juxtaposed *IDS* intron 7 and *IDS2* intron 7 *(25, 26)*. Because of the intragenic rearrangement, there would be more short sequenced reads connecting the distant genomic regions. Thus, the paired ends of the sequenced maternal DNA molecules that contained the recombinant would appear

to be as long as HMW (high molecular weight) DNA molecules when mapped to the reference genome. We used this feature to assign SNPs to the respective haplotypes, namely SNP alleles associated with the apparently long DNA molecules were assigned to the mutant-linked haplotype. The opposite SNP alleles were then assigned to the wild-type–linked haplotype (see online Supplemental Fig. 1). From RHDO analysis of maternal plasma DNA, there was an overrepresentation of mutant-linked SNP alleles and this indicated that the fetus had inherited the mutant allele from the mother. The result was concordant with the clinical diagnosis and the chorionic villi analysis.

Discussion

In this study, we used a direct haplotyping method to resolve the parental haplotypes across disease loci, which were then used to interpret targeted sequencing data obtained from maternal plasma DNA. Using this approach, the fetal mutation profiles in 12 of 13 families, at risk for a range of single gene diseases, were successfully deduced.

Our previous study showed that the entire fetal genome is present in maternal plasma in a constant relative proportion to maternal DNA *(13)*. It was therefore possible to noninvasively determine the fetal inheritance status in a genome-wide scale by combining parental haplotype information with massively parallel sequencing data of maternal plasma DNA. Since whole-genome haplotyping technologies were not mature in the past, haplotype information was derived from analyzing samples of related family members such as a proband *(14)*. However, this meant that for most practical purposes, the approach could only be applied to families where DNA from a previously affected member was available. With the use of linked-read sequencing for direct haplotyping, one can use the RHDO approach for noninvasive prenatal testing in families where no proband sample is available. Indeed, here we showed that haplotyping of the parental DNA was achieved for all 13 families. We showed that this direct whole-genome haplotyping method circumvented the need to analyze samples from related family members affected with the disease. This new development not only means that the cost of the analysis has reduced, it also means that noninvasive fetal genotyping could potentially be applied to most at-risk pregnancies.

Another advantage of haplotype-based methods over direct mutational analysis is that one could infer the fetal inheritance through quantitative assessment of informative SNP alleles in maternal plasma, obviating the need for tailor-made mutation-specific assays *(22, 23)*. Such tailor-made assays need to be optimized in good time to meet the requirements for a clinically acceptable turnaround time during pregnancy. Sometimes, mutation-specific assays cannot be as readily developed for some challenging genomic loci (e.g., repetitive regions, existence of homologous genes) or for certain mutations (deletions, inversions, gene recombinants). *CYP21A2* is one such example. The sequences of *CYP21A2* share high homology with the pseudogene cytochrome P450 family 21 subfamily A member 1, pseudogene (*CYP21A1P*). Because the fetal genotype was inferred from the SNP allelic ratios in maternal plasma, assays tailor-made for the *CYP21A2* mutations were not needed.

Instead, a series of probes for the target capture of SNPs surrounding of a group of clinically important single gene disease loci could be prestocked in the laboratory. The scale of the testing could be varied depending on clinical needs. For example, one may elect to use only target capture probes designed for the assessment of 1 disease locus at a time. This strategy is suitable for the assessment of high risk pregnancies either with a family history for a specific single gene disease or had been identified to be mutation carriers through screening programs *(27)*. Alternatively, target capture probes relevant for several disease loci could be pooled and be analyzed concurrently. This alternative strategy is useful when there are a number of gene loci to be tested, such as for the purpose of investigating fetal abnormalities, like congenital cardiac defects, detected by ultrasonography.

There is also the potential to apply this noninvasive testing approach in the public health setting aimed at the prenatal management of diseases that are of high prevalence in the community, for example cystic fibrosis, sickle cell anemia or the thalassemias, or diseases that would benefit from prenatal *(28)* or early neonatal treatment. When used as a public health screening tool, the capture probes are first used for carrier identification *(29)* where the linked-read sequencing of the parental DNA is used to determine the parental mutations and haplotype structures. The same probes are then used for the target capture of maternal plasma DNA for haplotype-based fetal genotype assessment. Thus, the workflow for the prenatal screening and detection of single gene diseases would be much more streamlined.

For the haplotype-based approach to be successful, some requirements must be fulfilled. Noninvasive deduction of the fetal genotype can only be achieved if the maternal plasma DNA data surrounding the disease locus are adequate to allow statistically significant dosage assessments between the parental haplotypes. The amount of sequence information needed is dependent on the fetal DNA fraction, the number of informative SNPs and the sequencing depth. For example, we could not provide fetal inheritance assessment for family H because of the inadequate number of informative SNPs as well as low plasma DNA sequence coverage of the highly repetitive region. Perhaps, the

capture probes targeting this locus could be redesigned to capture more SNPs. Computational simulation showed that if the number of SNPs reached 1000 with 200-fold sequencing depth, a statistical confident RHDO classifications can be generated even with low fractional fetal DNA concentrations *(14)*. In this study, we can classify a sample of fractional fetal DNA concentration as low as 4.7%. One should also be mindful of the effect of recombination as detected in 3 cases in this study. A recombination event may result in incorrect fetal genotype classification if it occurs at a genomic location near the mutation.

The protocol described in this study can readily be employed to many cases and the turnaround time required is about 1–2 weeks. We have demonstrated that the approach was applicable to a variety of single gene diseases. We believe that we have developed an approach that could be universally applied as a generic protocol for the noninvasive assessment of fetal single gene diseases. This development could make noninvasive prenatal assessment of fetal single gene diseases more widely adopted.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

Employment or Leadership: K.C.A. Chan, Xcelom and Cirina; Y.M.D. Lo, *Clinical Chemistry*, AACC, Xcelom, and Cirina; R.W.K. Chiu, Xcelom and Cirina.
Consultant or Advisory Role: K.C.A. Chan, Xcelom; R.W.K. Chiu, Xcelom; P. Jiang, Cirina.
Stock Ownership: K.C.A. Chan, Xcelom and Cirina; R.W.K. Chiu, Xcelom and Cirina; Y.M.D. Lo, Xcelom and Cirina.
Honoraria: R.W.K. Chiu, Illumina.
Research Funding: Research Grants Council of the Hong Kong SAR Government under the Theme-based research scheme (T12-403/15-N); Y.M.D. Lo, Hong Kong Research Grants Council Theme-Based Research Scheme T12-403/15-N, Sequenom, Sponsored Research Agreement.
Expert Testimony: None declared.
Patents: K.C.A. Chan, US patent 8 467 976; Y.M.D. Lo, US patent 8 467 976; R.W.K. Chiu, US patent 8 467 976.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, and final approval of manuscript.

Acknowledgments: We thank Dr. Liz Yuen for helpful discussion. We also thank Yongjie Jin for technical assistance.

References

1. Lo YM, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CW, Wainscoat JS. Presence of fetal DNA in maternal plasma and serum. Lancet 1997;350:485–7.
2. Chiu RW, Akolekar R, Zheng YW, Leung TY, Sun H, Chan KC, et al. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. BMJ 2011;342:c7401.
3. McCullough RM, Almasri EA, Guan X, Geis JA, Hicks SC, Mazloom AR, et al. Non-invasive prenatal chromosomal aneuploidy testing–clinical experience: 100 000 clinical samples. PLoS One 2014;9:e109173.
4. Lun FM, Chiu RW, Chan KC, Leung TY, Lau TK, Lo YM. Microfluidics digital PCR reveals a higher than expected fraction of fetal DNA in maternal plasma. Clin Chem 2008;54:1664–72.
5. Lo YM, Tein MS, Lau TK, Haines CJ, Leung TN, Poon PM, et al. Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. Am J Hum Genet 1998;62:768–75.
6. Costa JM, Benachi A, Gautier E. New strategy for prenatal diagnosis of X-linked disorders. N Engl J Med 2002; 346:1502.
7. Bustamante-Aragones A, Rodriguez de Alba M, Gonzalez-Gonzalez C, Trujillo-Tiebas MJ, Diego-Alvarez D, Vallespin E, et al. Foetal sex determination in maternal blood from the seventh week of gestation and its role in diagnosing haemophilia in the foetuses of female carriers. Haemophilia 2008;14:593–8.
8. Lo YM, Hjelm NM, Fidler C, Sargent IL, Murphy MF, Chamberlain PF, et al. Prenatal diagnosis of fetal RhD status by molecular analysis of maternal plasma. N Engl J Med 1998;339:1734–8.
9. Saito H, Sekizawa A, Morimoto T, Suzuki M, Yanaihara T. Prenatal DNA diagnosis of a single-gene disorder from maternal plasma. Lancet 2000;356:1170.
10. Chiu RW, Lau TK, Leung TN, Chow KC, Chui DH, Lo YM. Prenatal exclusion of beta thalassaemia major by examination of maternal plasma. Lancet 2002;360:998–1000.
11. Lun FM, Tsui NB, Chan KC, Leung TY, Lau TK, Charoenkwan P, et al. Noninvasive prenatal diagnosis of monogenic diseases by digital size selection and relative mutation dosage on DNA in maternal plasma. Proc Natl Acad Sci 2008;105:19920–5.
12. Tsui NB, Kadir RA, Chan KC, Chi C, Mellars G, Tuddenham EG, et al. Noninvasive prenatal diagnosis of hemophilia by microfluidics digital PCR analysis of maternal plasma DNA. Blood 2011;117:3684–91.
13. Lo YM, Chan KC, Sun H, Chen EZ, Jiang P, Lun FM, et al. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. Sci Transl Med 2010;2:61ra91.
14. New MI, Tong YK, Yuen T, Jiang P, Pina C, Chan KC, et al. Noninvasive prenatal diagnosis of congenital adrenal hyperplasia using cell-free fetal DNA in maternal plasma. J Clin Endocrinol Metab 2014;99: E1022–30.
15. Lam KW, Jiang P, Liao GJ, Chan KC, Leung TY, Chiu RW, Lo YM. Noninvasive prenatal diagnosis of monogenic diseases by targeted massively parallel sequencing of maternal plasma: application to beta-thalassemia. Clin Chem 2012;58:1467–75.
16. Zeevi DA, Altarescu G, Weinberg-Shukron A, Zahdeh F, Dinur T, Chicco G, et al. Proof-of-principle rapid noninvasive prenatal diagnosis of autosomal recessive founder mutations. J Clin Invest 2015;125:3757–65.
17. Kitzman JO, Mackenzie AP, Adey A, Hiatt JB, Patwardhan RP, Sudmant PH, et al. Haplotype-resolved genome sequencing of a Gujarati Indian individual. Nat Biotechnol 2011;29:59–63.
18. Amini S, Pushkarev D, Christiansen L, Kostem E, Royce T, Turk C, et al. Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing. Nat Genet 2014;46:1343–9.
19. Selvaraj S, R Dixon J, Bansal V, Ren B. Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. Nat Biotechnol 2013;31:1111–8.
20. Snyder MW, Adey A, Kitzman JO, Shendure J. Haplotype-resolved genome sequencing: experimental methods and applications. Nat Rev Genet 2015;16: 344–58.
21. Zheng GX, Lau BT, Schnall-Levin M, Jarosz M, Bell JM, Hindson CM, et al. Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. Nat Biotechnol 2016;34:303–11.
22. Lench N, Barrett A, Fielding S, McKay F, Hill M, Jenkins L, et al. The clinical implementation of non-invasive prenatal diagnosis for single-gene disorders: challenges and progress made. Prenat Diagn 2013;33: 555–62.
23. Verhoef TI, Hill M, Drury S, Mason S, Jenkins L, Morris S, Chitty LS. Non-invasive prenatal diagnosis (NIPD) for single gene disorders: cost analysis of NIPD and invasive testing pathways. Prenat Diagn 2016;36:636–42.
24. Chiu RW, Poon LL, Lau TK, Leung TN, Wong EM, Lo YM. Effects of blood-processing protocols on fetal and total DNA quantification in maternal plasma. Clin Chem

2001;47:1607–13.
25. Lualdi S, Regis S, Di Rocco M, Corsolini F, Stroppiano M, Antuzzi D, Filocamo M. Characterization of iduronate-2-sulfatase gene-pseudogene recombinations in eight patients with mucopolysaccharidosis type II revealed by a rapid PCR-based method. Hum Mutat 2005;25:491–7.
26. Bondeson ML, Dahl N, Malmgren H, Kleijer WJ, Tonnesen T, Carlberg BM, Pettersson U. Inversion of the IDS gene resulting from recombination with IDS-related sequences is a common cause of the Hunter syndrome. Hum Mol Genet 1995;4:615–21.
27. Samavat A, Modell B. Iranian national thalassaemia screening programme. BMJ 2004;329:1134–7.
28. New MI, Abraham M, Yuen T, Lekarev O. An update on prenatal diagnosis and treatment of congenital adrenal hyperplasia. Semin Reprod Med 2012;30:396–9.
29. Bell CJ, Dinwiddie DL, Miller NA, Hateley SL, Ganusova EE, Mudge J, et al. Carrier testing for severe childhood recessive diseases by next-generation sequencing. Sci Transl Med 2011;3:65ra4.