



Noninvasive prenatal test of single-gene disorders by linked-read direct haplotyping: application in various diseases

Jee-Soo Lee¹ · Kyung Bok Lee² · Han Song³ · ChoongHyun Sun³ · Man Jin Kim¹ · Sung Im Cho¹ · Young Kyung Lee⁴ · Sung Sup Park¹ · Moon-Woo Seong¹

Received: 4 February 2020 / Revised: 26 August 2020 / Accepted: 20 October 2020 / Published online: 24 November 2020
© The Author(s), under exclusive licence to European Society of Human Genetics 2020

Abstract

Direct haplotyping enables noninvasive prenatal testing (NIPT) without analyzing proband, which is a promising strategy for pregnancies at risk of an inherited single-gene disorder. Here, we aimed to expand the scope of single-gene disorders that NIPT using linked-read direct haplotyping would be applicable to. Three families at risk of myotonic dystrophy type 1, lipoid congenital adrenal hyperplasia, and Fukuyama congenital muscular dystrophy were recruited. All cases exhibited distinct characteristics that are often encountered as hurdles (i.e., repeat expansion, identical variants in both parents, and novel variants with retrotransposon insertion) in the universal clinical application of NIPT. Direct haplotyping of parental genomes was performed by linked-read sequencing, combined with allele-specific PCR, if necessary. Target *DMPK*, *STAR*, and *FKTN* genes in the maternal plasma DNA were sequenced. Posterior risk calculations and an Anderson–Darling test were performed to deduce the maternal and paternal inheritance, respectively. In all cases, we could predict the inheritance of maternal mutant allele with > 99.9% confidence, while paternal mutant alleles were not predicted to be inherited. Our study indicates that direct haplotyping and posterior risk calculation can be applied with subtle modifications to NIPT for the detection of an expanded range of diseases.

Introduction

Since the discovery of cell-free fetal DNA (cffDNA) in maternal blood about 20 years ago [1], advances have been made in noninvasive fetal genomic analysis, so-called noninvasive prenatal testing (NIPT). In parallel

with the development of next-generation sequencing (NGS), NIPT has been implemented in clinical prenatal care by several health care providers that mostly focus on screening of fetal chromosomal aneuploidy, including trisomy 21, trisomy 18, trisomy 13, and monosomy X [2–4].

Single-gene disorders are another context in which NIPT might be applied, particularly for couples with fetus at risk of an inherited monogenic disease [5]. Different NIPT strategies need to be devised based on the inheritance pattern of the specific disease (e.g., autosomal dominant, autosomal recessive, and X-linked) and according to whether the paternal or maternal inheritance should be assessed. NIPT of single-gene disorders includes two major steps: 1) reconstruction of parental haplotypes and 2) maternal plasma sequencing followed by deduction of the fetal genotype [6].

The parental haplotypes have been frequently reconstructed through inferential approaches based on the genotypes of mother, father, and the proband, which is not a trivial task. This method is more challenging when the genotype of the proband is unavailable [5, 7]. However, recently, a microfluidics-based linked-read sequencing

Supplementary information The online version of this article (<https://doi.org/10.1038/s41431-020-00759-9>) contains supplementary material, which is available to authorized users.

✉ Moon-Woo Seong
mwseong@snu.ac.kr

- ¹ Department of Laboratory Medicine, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Republic of Korea
- ² Department of Statistics, Seoul National University, Seoul, Republic of Korea
- ³ Genome Opinion Inc., 7, Yeonmujang 5ga-gil, Seongdong-gu, Seoul, Republic of Korea
- ⁴ Department of Laboratory Medicine, Hallym University Sacred Heart Hospital, Anyang, Republic of Korea

technology (10x genomics) was reported, which enables direct determination of haplotypes, thereby resolving the drawbacks of indirect haplotyping mentioned above [8]. This direct haplotyping has been successfully applied to several monogenic disorders, namely, congenital adrenal hyperplasia (CAH), Ellis-van Creveld syndrome, hemophilia, and Hunter syndrome, as reported by Hui et al. [7]. However, this approach remains at the proof-of-concept stage and warrants further examination to establish its clinical validity. If this approach could be broadly adaptable to a range of monogenic diseases, NIPT of single-gene disorders could be more easily carried out in clinical practice.

In this study, we aimed to expand the range of NIPT for detecting single-gene disorders using linked-read direct haplotyping. We adapted, with a subtle modification, the linked-read direct phasing described by Hui et al. to three families at risk of Myotonic dystrophy type 1 (DM1), lipoid CAH, and Fukuyama congenital muscular dystrophy (FCMD). In the case of DM1, we integrated the allele-specific polymerase chain reaction (AS-PCR) with linked-read direct haplotyping to efficiently target expanded CTG repeat sequences. To determine the applicability of different statistical approaches, we integrated relative haplotype dosage (RHDO) analysis and posterior risk calculation. The overall strategy for noninvasive diagnosis used in this study is shown in Fig. 1.

Materials and methods

Sample collection and preparation

Peripheral blood samples were collected from pregnant women and their partners with informed consent, in parallel with invasive procedures during pregnancy. These paternal and maternal blood samples were centrifuged at $1600 \times g$ for 10 min at room temperature. For maternal blood samples, the plasma fractions were sequentially centrifuged at $13,000 \times g$ for 10 min at room temperature. Genomic DNA (gDNA) was extracted from the buffy coat with Gentra Puregene Blood kit (Gentra Systems, Minneapolis, MN, USA), according to the manufacturer's instructions. Plasma DNA (i.e., cell-free DNA; cfDNA) was manually extracted from 1 mL of maternal plasma with a QIAmp circulating nucleic acid kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. The cfDNA was eluted in 50 μ L Buffer AVE provided by Qiagen and stored at 4 °C until further analysis. This study was approved by the institutional review board of the Seoul National University Hospital.

Direct haplotyping

We reconstructed multiple haplotype phase blocks directly from a single sample, through the 10x Genomics

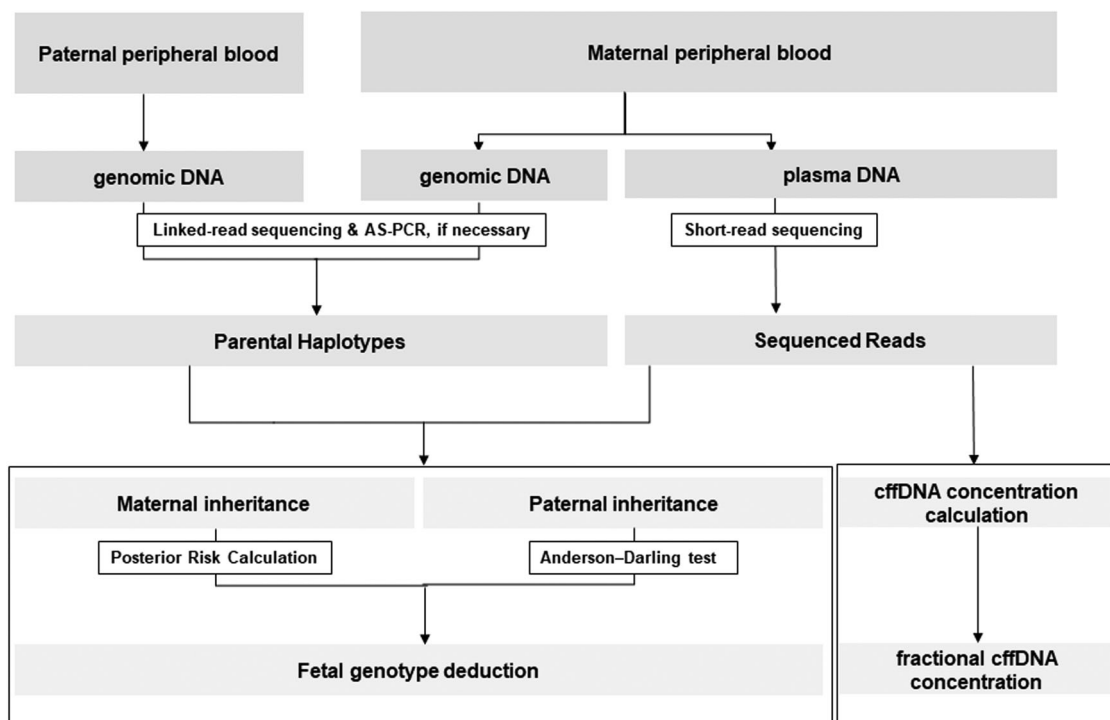


Fig. 1 Flow chart of the overall strategy for noninvasive prenatal diagnosis of single-gene disorders in this study. AS-PCR allele-specific PCR, RHDO relative haplotype dosage, cffDNA cell-free fetal DNA.

linked-read sequencing (10x Genomics, Pleasanton, CA, USA). Briefly, 1 ng of gDNA was loaded with reagents from Chromium™ Genome Reagent Kits v2 (10x Genomics). A microfluidic device was used to partition each high molecular weight (HMW) DNA in 10x™ barcoded gel beads, which enabled gDNA to be tagged with unique 16 bp barcodes, into ~1.5 million oil droplets in emulsion (i.e., *GEM*). Target enrichment for the linked-read libraries was then conducted using the Agilent SureSelectXT Custom kit [targeting 3.2 Mb, including *DMPK* (NM_001081563.2), *STAR* (NM_000349.2), and *FKTN* (NM_001079802.1) genes], according to the recommended protocols (Agilent, Santa Clara, CA, USA). Libraries were sequenced on the Illumina HiSeq™ 2500 platform (Illumina, San Diego, CA, USA).

Phasing of linked sequence reads was performed with Long Ranger software (version 2.1.6; <https://github.com/10XGenomics/longranger>) from 10x Genomics (Pleasanton, CA, USA). The barcoded reads were aligned to the GRCh37/hg19 using Lariat aligner, which uses Burrows–Wheeler aligner to map the reads to the reference genome [9]. The code is available at <https://github.com/10XGenomics/lariat> [10]. Variant calling was performed using FreeBayes version 0.9.21. Linked reads came from the same HMW DNA molecule; thus, variants in the same barcoded reads were strung together into haplotype blocks. All the sequencing data obtained in this study are submitted to the Sequence Read Archive (SRA) (<https://www.ncbi.nlm.nih.gov/sra>) under accession number PRJNA644213.

Allele-specific PCR for disease-associated haplotypes

In the family at risk for a fetus with DM1 (family A), the mother was heterozygous for expanded CTG allele with 79 repeats in the 3′ untranslated region (3′ UTR) of *DMPK* gene [i.e., NC_000019.9:g.(46273386_46273557)ins(135)]. We additionally designed a strategy to identify a haplotype associated with the repeat expansion in maternal gDNA, using AS-PCR. We selected an informative SNP (rs635299) flanking the causative CTG repeats, and where the mother was heterozygous (Hap I, A/ Hap II, C). The AS-PCR for *DMPK* genotyping was carried out using allele-specific forward primers for both allelic variants of the SNP position (Supplemental Table 1).

To ascertain if these allele-specific primers efficiently discriminate two different alleles, we carried out AS-PCR experiment using two normal control DNA samples, A/A and C/C, respectively, at rs635299. To determine the allele associated with the repeat expansion, we then performed AS-PCR with maternal gDNA. Reactions were performed using the following annealing conditions, extension time, and number of cycles: 55 °C for 30 s; 60 s; 35 cycles. Then, we performed 2% agarose gel electrophoresis to identify the

amplified PCR product. Each experiment was performed in duplicate.

Maternal plasma DNA sequencing

For maternal plasma DNA, 10 ng DNA was used for library construction without additional fragmentation. Libraries were constructed using the SureSelectXT Custom kit (targeting 3.2 Mb, including *DMPK*, *STAR*, and *FKTN* genes), according to the recommended protocols (Agilent). After end-repair, adenylation at the 3′ end, and ligation with adapters, plasma DNA libraries were enriched with target capture probes. Libraries were sequenced on Illumina HiSeq™ 2500 platform (Illumina). On average, 117 million reads of 150-bp paired-end sequencing were obtained from each sample. FASTQ files from the sequencing results were aligned to the GRCh37/hg19, and the variants were called using NextGene software version v2.4.2.2 (SoftGenetics, LLC, State College, PA, USA).

Fractional fetal concentration

The fraction of fetal DNA concentration was estimated according to the methods of Lo et al. [11]. SNPs were selected where both mother and father were homozygous, but each had different genotypes; the fetal genotype was predicted to be heterozygous. Two times the paternal specific allele count was divided by the total count of both alleles. SNPs for measuring the fraction of fetal DNA concentration were selected from regions for target enrichment in each family.

Fetal genotype deduction

Maternal inheritance

We integrated RHDO analysis and posterior risk calculation [11, 12]. Informative SNPs were selected where the mother was heterozygous, and the father was homozygous for one of the alleles. Each SNP was further classified as either α SNP or β SNP: α SNP was defined when the homozygous paternal alleles were identical to maternal Hap I (haplotype linked with the disease-causing variant). β SNP was defined when the paternal alleles were identical to maternal Hap II (haplotype linked with wild-type allele) [7]. Then, we calculated the overdispersion in sequence data, and the posterior risk of a maternal mutant allele being inherited by the fetus according to the methods of Vermeulen et al. [12]. It has been assumed that the reads of the two types of SNPs (α and β SNPs) were independent. The number of informative reads theoretically required to determine overrepresentation of mutant-linked haplotype in maternal plasma was computed as described by Vermeulen et al. [12].

Paternal inheritance

For the families whose father carried a mutant allele, we attempted to determine if the paternal mutant allele was inherited. In these cases, informative SNPs were selected where the mother was homozygous, and the father was heterozygous for one of the alleles. The paternal-specific SNP alleles detected in maternal plasma were assessed for belonging to paternal mutant-linked haplotype or wild-linked haplotype. Anderson–Darling (AD) test was performed to determine the significant difference in sequence read counts between paternal mutant-linked haplotype and wild-linked haplotype.

Fetal DNA genotyping

Fetal DNA was extracted from amniotic fluid or chorionic villus sample with the Gentra Puregene DNA isolation kit (Gentra Systems, Minneapolis, MN, USA) according to the manufacturer's instructions. To confirm the fetal genotype at additional genomic positions (Supplementary Table 2), genomic regions flanking the target SNPs were amplified. The amplified product was directly sequenced with an ABI

PRISM 3730xl DNA Analyzer (Applied Biosystems, Foster City, CA, USA) using the BigDye Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems). We compared the fetal DNA genotyping results to the deduced fetal genotypes.

Results

Clinical characteristics of the cases

We analyzed three families having a risk for DM1 (family A), Lipoid CAH (family B), and FCMD (family C), respectively. The pedigrees and clinical information of the recruited families are shown in Fig. 2 and Table 1. Plasma samples in each family were collected during the first trimester (11 weeks and 5 days gestation in family A, 12 weeks and 4 days gestation in family B, and 11 weeks and 0 days gestation in family C).

In family A, the mother was heterozygous for a normal allele of 13 CTG repeats, and an expanded allele of 79 CTG repeats in the 3' UTR of *DMPK* [i.e., NC_000019.9:g.(46273386_46273557)ins(135)], while the father carried

Fig. 2 Pedigree of three families with DM1 (family A), Lipoid CAH (family B), and FCMD (family C). The size of their CTG repeats is marked in the pedigree of the family A. Proband is marked with arrows. DM1 Myotonic dystrophy type 1, Lipoid CAH lipoid congenital adrenal hyperplasia, FCMD Fukuyama congenital muscular dystrophy.

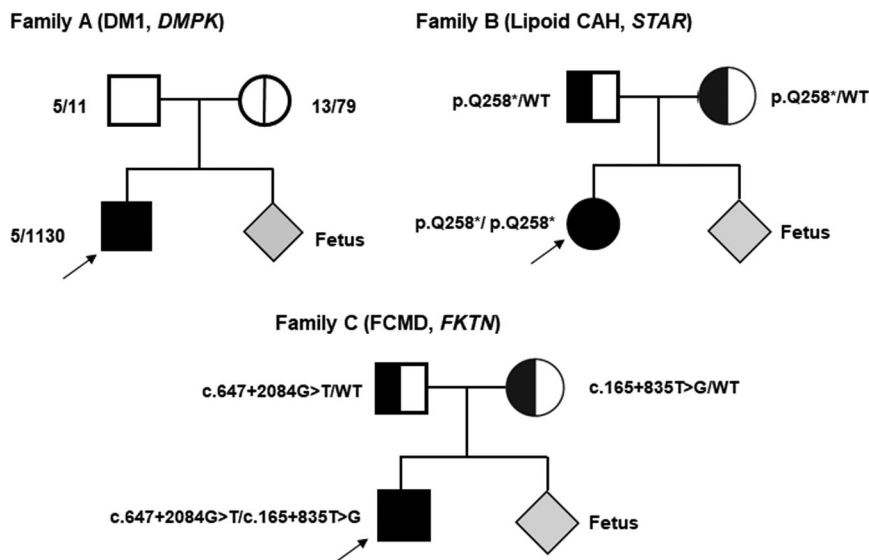


Table 1 Clinical information of recruited families.

Family	Diseases	Gene	Pattern of inheritance	Location of gene	Maternal genotype	Paternal genotype	Gestation age	Fetal DNA concentration (%)
A	DM1	<i>DMPK</i>	AD	chr19	CTG repeat size, 13/79	CTG repeat size, 5/11	11 + 5	12.6
B	Lipoid CAH	<i>STAR</i>	AR	chr8	c.772C>T	c.772C>T	12 + 4	16.3
C	FCMD	<i>FKTN</i>	AR	chr9	c.165+835T>G	c.647+2084G>T	11 + 0	12.4

DM1 Myotonic dystrophy type 1, Lipoid CAH lipoid congenital adrenal hyperplasia, FCMD Fukuyama congenital muscular dystrophy, AD autosomal dominant, AR autosomal recessive.

non-expanded normal CTG repeat alleles (repeat size, 5/11). Our analysis indicated that the fetus inherited a CTG expanded allele from the mother. In family B, both parents were carriers of a heterozygous variant in *STAR* c.772C>T, p.(Gln258*) that affects function, and our results indicated that the fetus had inherited maternal mutant allele along with paternal wild-type allele. In family C, the mother carried a heterozygous *FKTN* deep-intronic variant NG_008754.1 (NM_001079802.1): c.165+835T>G that was previously reported as a novel intronic marker for retrotransposon (RT) insertion variant [13]. The father carried a heterozygous deep-intronic *FKTN* variant (c.647+2084G>T) that was recently identified to cause abnormal splicing and a recessive form of congenital muscular dystrophy in Korean families [14]. Our results revealed that the fetus did not inherit the paternal mutant allele, but a maternal variant allele of c.165+835T>G. Concomitantly, these findings were confirmed by conventional prenatal diagnosis (i.e., amniocentesis and chorionic villus sampling) in each family.

Sequencing

For parental gDNA from each family, the linked-read library was prepared and sequenced to mean coverage of 593-fold (range, 537–675). The linked-read sequencing data are summarized in Supplemental Table 3. The mean molecule length was 33 kb, and the N50 phase block length averaged 632 kb (range 219 kb–1.4 Mb). These phasing results indicated that linked-read sequencing provided long-range genomic contiguity suitable for consequent analysis. For maternal plasma DNA from each family, the average coverage of the target region was 230-fold (range, 145–375) with 99.9% of target region covered by at least 10 reads. The summary of maternal DNA sequencing analysis is presented in Supplemental Table 4. The fetal DNA fraction in the maternal plasma was estimated to be 12.6% (family A; 11 wk + 5 d), 16.3% (family B; 12 wk + 4 d), and 12.4% (family C; 11 wk + 0 d), respectively (Table 1).

Noninvasive assessment for repeat expansion disease (family A)

DM1 is an autosomal dominant disorder associated with the expansion of a CTG repeat, and we attempted to determine whether the fetus inherited the maternal haplotype linked with expanded CTG repeats. We noted that it is challenging to directly phase the allele with the CTG expansions where the expanded allele size was greater than the length of short-read sequencing-derived reads. Thus, we applied a different strategy to this family; we combined AS-PCR with linked-read phasing.

We selected the informative SNP (rs635299, A/C) flanking the causative CTG repeat and where the mother was heterozygous (Hap I, A/ Hap II, C), and assessed which SNP allele (A or C) belongs to a maternal expanded allele. Using the mother's gDNA, AS-PCR was performed with primers specific for each of the two alleles at rs635299, and the PCR-amplified products were loaded on an agarose gel. The PCR product amplified with the A allele-specific primer (Hap I) was much larger than that amplified with the C allele-specific primer (Hap II) (Supplemental Fig. 1). The PCR results indicated that the size of the expanded alleles corresponded to approximately 1000 bp, while normal allele yielded a shorter fragment of 750 bp with the C allele-specific primer. These results demonstrated that the expanded CTG repeats of the mother belonged to Hap I and not Hap II.

We then calculated the posterior risk of maternal mutant-linked haplotype being transmitted (Table 2). The number of phased informative SNPs was 1796 (mother) and 1794 (father) for this family, and the number of sequence reads exceeded the theoretically required number of reads. A total of 264 α SNPs and 167 β SNPs were used for posterior risk calculation. Under the assumption that maternal Hap I is inherited, the fraction of Hap I in maternal plasma was estimated as 56.3% (α SNP) and 50.0% (β SNP), respectively. Under the assumption that maternal Hap II is inherited, the fraction of Hap I in maternal plasma was

Table 2 Fetal genotype deduction using informative SNPs and maternal plasma DNA sequencing data.

Family	Fetal DNA (%)	Theoretical requirements*	Reads**	Maternal inheritance			Paternal inheritance	
				α SNPs	β SNPs	Posterior risk***	No. of SNPs used	<i>P</i> value
A	12.6	1991	56,859	264	167	>99.9%	Not tested	Not tested
B	16.3	1143	16,147	118	6	>99.9%	223	8.8×10^{-10}
C	12.4	2060	130,555	300	146	>99.9%	94	2.0×10^{-21}

*Number of reads theoretically required to reliably detect a dosage imbalance of maternal haplotype for a given fractional fetal concentration in each family.

**Number of sequence reads of informative SNPs used for calculating posterior risk.

***Posterior risk for inheritance of maternal haplotype linked with mutation.

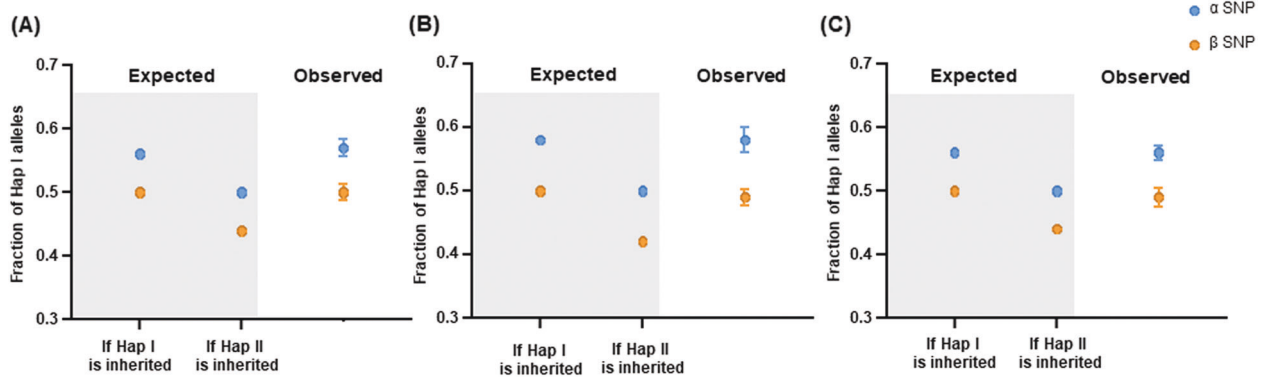


Fig. 3 Distribution of the fraction of Hap I alleles according to the inheritance of either maternal Hap I or Hap II in three families with DM1 (A), Lipoid CAH (B), and FCMD (C). The expected fraction (in the gray box) under each category was compared with the observed distribution, for α SNP (blue) and β SNP (orange),

estimated as 50.0% (α SNP) and 43.70% (β SNP), respectively (Supplemental Table 5). The observed distributions of Hap I in maternal plasma were found to be much closer to the expected fraction under the assumption that maternal Hap I is inherited (α SNP, mean 57.04%, β SNP, mean 49.66%) (Fig. 3A). Through statistical analysis of comparing the expected distributions with observed distributions of the allele fraction, we were able to compute the posterior risk for maternal Hap I being transmitted with >99.9% probability (Table 2). Additional fetal DNA genotyping showed that genotype of the allele inherited from the mother was identical to that of maternal Hap I (Supplementary Table 2).

Noninvasive assessment for autosomal recessive diseases (family B and C)

Given that both parents in family B (Lipoid CAH) and C (FCMD) were heterozygous carriers of autosomal recessive disease, fetal inheritance from each parent had to be deduced with a stepwise approach in these two families.

We first computed the posterior risk of maternal mutant-linked haplotype being transmitted (Table 2). The number of phased informative SNPs was 1306 (mother) and 1296 (father) for family B, and 2016 (mother) and 2020 (father) for family C. In a posterior risk calculation, 124 SNPs (118 α SNPs and 6 β SNPs) and 446 SNPs (300 α SNPs and 146 β SNPs) were used for family B and C, respectively. We calculated the expected fraction of the Hap I alleles in maternal plasma, for a given fetal fractional concentration, and compared it with the observed fraction under each assumption that maternal Hap I is inherited or maternal Hap II is inherited (Supplemental Table 4, Fig. 3B, C). Under the assumption that maternal Hap I is inherited, the Hap I fraction in maternal plasma was estimated as 58.15% (α SNP) and 50.0% (β SNP) for family B, and as 56.20% (α SNP) and 50.0% (β SNP) for

family C. Under the assumption that maternal Hap II is inherited, the Hap I fraction in maternal plasma was estimated as 50.0% (α SNP) and 41.85% (β SNP) for family B, and as 50.0% (α SNP) and 43.80% (β SNP) for family C. In both families, the observed distributions of Hap I in maternal plasma were found to be much closer to the expected fraction under the assumption that maternal Hap I was inherited (family A, α SNP, mean 58.23%, β SNP, mean 48.80%; family B, α SNP, mean 56.01%, β SNP, mean 48.89%) (Fig. 3B, C). Through statistical analysis, we were able to determine the posterior risk for maternal Hap I being transmitted with >99.9% probability in both families (Table 2). These findings were confirmed by fetal DNA genotyping (Supplementary Table 2).

To deduce the inheritance of the paternal variant, we detected paternal specific SNP alleles in maternal plasma. The informative SNP loci were where the father was heterozygous, and the mother was homozygous for one of the alleles: family A had 223 informative SNPs encompassing the captured *STAR* region, and family B had 94 informative SNPs encompassing the captured *FKTN* region. Each AD test indicated statistically significant difference of allelic counts between the two groups of paternal-unique reads (mutant linked vs. wild-type linked, P -value = 8.8×10^{-10} [family B], P -value = 2.0×10^{-21} [family C]), which suggested that maternal plasma contains more paternal-specific SNP alleles belonging to wild-type-linked haplotype than paternal-specific SNP alleles belonging to the mutant-linked haplotype (Table 2). Therefore, we identified that paternal mutant alleles were not transmitted to both fetuses in family B and C.

Therefore, we identified that paternal mutant alleles were not transmitted to both fetuses in family B and C.

Discussion

In this study, we applied noninvasive diagnostics to DM1, Lipoid CAH, and FCMD through a stepwise approach:

linked-read sequencing and AS-PCR for parental haplotyping, and targeted sequencing of maternal plasma DNA. Notably, we estimated posterior risk of a maternal mutant allele being transmitted to the fetus, which demonstrated that diverse statistical approaches are applicable to a range of single-gene disorders. In clinical laboratories, gDNA from a proband is often unavailable, and both parents are often carriers of the same variant; NIPT using linked-read sequencing could compensate for these drawbacks and provide benefits to more families at risk for single-gene disorders.

Three recent NIPT studies published in the field of monogenic disorders have demonstrated the integration of direct haplotyping (i.e., linked-read sequencing, or targeted locus amplification) and targeted maternal plasma DNA sequencing, although in a limited range of single-gene disorders (e.g., cystic fibrosis, CAH, β -thalassemia, hemophilia, and DMD) [7, 12, 15]. The current study is, to our knowledge, the first to show the feasibility of direct haplotyping-based NIPT in DM1, Lipoid CAH, and FCMD, with a slight modification. In addition, a major strength of this study is that the inferences were evaluated by validation analyses, which highlighted the reliability of this method in terms of accurate assessment.

In family A, at risk of CTG repeat expansion, AS-PCR was used to identify which of the haplotypes was linked with the repeat expansion. The heterozygote SNP rs635299 served as an anchor to be phased to flanked region. We therefore successfully linked the CTG repeat expansion to pre-constructed maternal haplotypes. We speculated that our AS-PCR & linked-read-based strategy could discriminate expanded allele from wild-type alleles and allow for the phasing of haplotypes in the families with various types of repeat expansion.

In family B, at risk of lipoid CAH, both parents were carriers of the same single nucleotide *STAR* c.772C>T p. (Gln258*) variant. This variant that causes the deletion of the 28 C-terminal amino acids of the StAR protein accounts for 80% among the known mutant alleles in Korean and Japanese populations [16]. This variant was detected more frequently (92.3%) in Korean mutant alleles [17]. For this reason, the linked-read-based strategy that generates reads integrated with a barcode and then traces the reads back to the individual DNA molecule would provide benefit to couples carrying identical hotspot variant in a certain gene [18].

In family C, at risk of FCMD, the proband had compound heterozygous variants, c.[165+835T>G];[647+2084G>T][13, 14]. Previously, Lim et al. described that *FKTN* c.165+835T>G always occurs in the same phase with an RT insertion in the FCMD cohort [13]; therefore, *FKTN* c.165+835T>G variant is an intronic marker for the RT insertion variant. To enhance the applicability of noninvasive diagnosis in clinical laboratory, we combined this

intronic marker in maternal gDNA with direct linked-read sequencing, which allows us to successfully reconstruct the haplotypes.

The major limitation of our study was the small number of families analyzed. However, the families analyzed in this study showed distinct characteristics that could be often encountered as hurdles (i.e., repeat expansion, identical variants in both parents, and novel variants with RT insertion) in the universal application of NIPT in clinical laboratories. Further studies using additional families could provide more reliable evidence for the wide application of our approach in single-gene disorders. Another limitation is that this approach is rather expensive and labor intensive to be used in routine diagnosis than alternative approaches: inheritance of paternal-specific alleles can be identified by directly confirming the presence of paternal specific variant in maternal plasma DNA. Also the use of droplet digital PCR recently reported by Camunas-Soler et al. does not require parental haplotyping, which can be directly applied to the maternal plasma DNA [5, 19].

In summary, we directly phased parental haplotypes using linked-read sequencing that combined with AS-PCR if necessary, and we estimated the posterior risk to successfully deduce fetal genotypes in families at risk for DM1, Lipoid CAH, and FCMD. Our results demonstrated that we could expand the range of diseases where noninvasive diagnosis could be applicable in clinical laboratories.

Acknowledgements

Funding This study was supported by grant 03-2016-0050 from the Seoul National University Hospital (SNUH) Research Fund.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

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

References

- Lo YM, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CW, et al. Presence of fetal DNA in maternal plasma and serum. *Lancet*. 1997;350:485–7.
- Hudecova I, Chiu RW. Non-invasive prenatal diagnosis of thalassemias using maternal plasma cell free DNA. *Best Pract Res Clin Obstet Gynaecol*. 2017;39:63–73.
- Wong FC, Lo YM. Prenatal diagnosis innovation: genome sequencing of maternal plasma. *Annu Rev Med*. 2016;67:419–32.
- Chitty LS, Lo YM. Noninvasive prenatal screening for genetic diseases using massively parallel sequencing of maternal plasma DNA. *Cold Spring Harb Perspect Med*. 2015;5:a023085.
- Chiu EKL, Hui WWI, Chiu RWK. cfDNA screening and diagnosis of monogenic disorders - where are we heading? *Prenat Diagn*. 2018;38:52–58.

6. Wong AI, Lo YM. Noninvasive fetal genomic, methylomic, and transcriptomic analyses using maternal plasma and clinical implications. *Trends Mol Med.* 2015;21:98–108.
7. Hui WW, Jiang P, Tong YK, Lee WS, Cheng YK, New MI, et al. Universal haplotype-based noninvasive prenatal testing for single gene diseases. *Clin Chem.* 2017;63:513–24.
8. 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.
9. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25:1754–60.
10. Bishara A, Liu Y, Weng Z, Kashef-Haghighi D, Newburger DE, West R, et al. Read clouds uncover variation in complex regions of the human genome. *Genome Res.* 2015;25:1570–80.
11. 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.
12. Vermeulen C, Geeven G, de Wit E, Verstegen M, Jansen RPM, van Kranenburg M, et al. Sensitive monogenic noninvasive prenatal diagnosis by targeted haplotyping. *Am J Hum Genet.* 2017;101:326–39.
13. Lim BC, Lee S, Shin JY, Hwang H, Kim KJ, Hwang YS, et al. Molecular diagnosis of congenital muscular dystrophies with defective glycosylation of alpha-dystroglycan using next-generation sequencing technology. *Neuromuscul Disord.* 2013;23:337–44.
14. Lim BC, Ki CS, Kim JW, Cho A, Kim MJ, Hwang H, et al. Fukutin mutations in congenital muscular dystrophies with defective glycosylation of dystroglycan in Korea. *Neuromuscul Disord.* 2010;20:524–30.
15. Jang SS, Lim BC, Yoo SK, Shin JY, Kim KJ, Seo JS, et al. Targeted linked-read sequencing for direct haplotype phasing of maternal DMD alleles: a practical and reliable method for non-invasive prenatal diagnosis. *Sci Rep.* 2018;8:8678.
16. Christenson LK, Strauss JF 3rd. Steroidogenic acute regulatory protein (StAR) and the intramitochondrial translocation of cholesterol. *Biochim Biophys Acta.* 2000;1529:175–87.
17. Kim CJ. Congenital lipoid adrenal hyperplasia. *Ann Pediatr Endocrinol Metab.* 2014;19:179–83.
18. Xia LC, Bell JM, Wood-Bouwens C, Chen JJ, Zhang NR, Ji HP. Identification of large rearrangements in cancer genomes with barcode linked reads. *Nucleic Acids Res.* 2018;46:e19.
19. Camunas-Soler J, Lee H, Hudgins L, Hintz SR, Blumenfeld YJ, El-Sayed YY, et al. Noninvasive prenatal diagnosis of single-gene disorders by use of droplet digital PCR. *Clin Chem.* 2018;64:336–45.