의학박사 학위논문

Noninvasive Prenatal Diagnosis of Duchenne Muscular Dystrophy: Comprehensive Genetic Diagnosis in Carrier, Proband, and Fetus

산모혈장의 태아 DNA를 이용한 듀센형 근이영양증 가계의 비침습적 산전유전자 진단법 개발

2017년 4월

서울대학교 대학원

의학과 뇌신경과학 전공

임 병 찬

ABSTRACT

BACKGROUND: Noninvasive prenatal diagnosis of monogenic disorders using maternal plasma and targeted massively parallel sequencing is being investigated actively. We demonstrated previously that comprehensive genetic diagnosis of a Duchenne muscular dystrophy (DMD) patient is feasible using a single targeted sequencing platform. In the present study, we demonstrate the applicability of this approach to carrier detection and noninvasive prenatal diagnosis.

METHODS: Custom solution-based target enrichment was designed to cover the entire *DMD* region. Targeted massively parallel sequencing was performed using genomic DNA from four mother and proband pairs to test whether carrier status could be detected reliably. Maternal plasma DNA at varying gestational weeks was collected from the same families and sequenced using the same targeted-platform to predict the inheritance of the *DMD* mutation by their fetus. Overrepresentation of an inherited allele was determined by comparing the allele fraction of two phased haplotypes after examining and correcting for the recombination event.

RESULTS: The carrier status of deletion/duplication and point mutations was detected reliably through using a single targeted massively parallel sequencing platform. Whether the fetus had inherited the *DMD* mutation was predicted correctly in all four families as early as 6 weeks and 5 days of gestation. In one of these, detection of the recombination event and reconstruction of the phased haplotype produced a correct diagnosis.

CONCLUSIONS: Noninvasive prenatal diagnosis of DMD is feasible using a single targeted massively parallel sequencing platform with tiling design.

Keywords: Noninvasive prenatal diagnosis, Duchenne muscular dystrophy, cell-free fetal DNA, massively parallel sequencing

Student number: 2009-30580

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

cffDNA – cell-free fetal DNA

DMD – Duchenne muscular dystrophy

SNV – single nucleotide variant

Indel – small insertion and deletion

CONTENTS

INTRODUCTION

Discovery of the presence of cell-free fetal DNA (cffDNA) in maternal plasma offers a powerful tool for the development of noninvasive prenatal genetic diagnosis (1). The application to prenatal diagnosis has been accelerated by the introduction of massively parallel sequencing technology (2, 3). Prenatal tests capable of detecting aneuploidies using cffDNA have been commercialized and are highly sensitive and accurate (4). Several studies have confirmed the accuracy of whole-genome sequencing and sequencing after target enrichment of cffDNA by demonstrating the relatively even distribution of fetal and maternal DNA across the entire genome (5–7). This result provides the basis for extending the applications to monogenic disorders, which comprise a larger proportion of genetic diseases than chromosomal aneuploidies.

However, unlike the rapid incorporation of aneuploidy detection into clinical practice, the application to monogenic disorders is far more complex and has many obstacles to overcome. Technically, the low and variable fraction of cffDNA in maternal plasma limits the reliable detection of fetal variants at the single-nucleotide level. Moreover, complex ethical and socioeconomic issues limit the implementation of noninvasive genome-wide screening in the prenatal diagnosis of monogenic disorders in pregnant women without a known increased risk. Therefore, for clinical applications, the ideal platform needs to be a targeted design that can ensure deep coverage and be equally applicable to the proband and carrier, and for prenatal diagnosis.

We have developed a method that allows the comprehensive genetic diagnosis of a Duchenne muscular dystrophy (DMD) patient. We have shown that this method is feasible when used with a targeted massively parallel sequencing platform (8). Targeting the entire exonic and intronic regions produced nearly continuous uniform coverage across the *DMD* gene,

enabling identification of both large deletions/duplications and point mutations. Because this method is sensitive enough to detect a dosage imbalance, the mother's carrier status could be easily identified with the same approach. In addition, because about 1,000 heterozygote sites could be used to analyze maternal X alleles at two phases, this method may be also applicable to prenatal diagnosis using cffDNA by detecting haplotype imbalances between two phased haplotypes in the *DMD* gene. This haplotype-based imbalance analysis via either whole-genome sequencing or targeted sequencing of maternal plasma DNA has been substantiated in models of several diseases, including beta-thalassemia, congenital adrenal hyperplasia, and congenital deafness (5, 9-11). Specifically, New et al. adopted a similar approach using targeted sequencing and a tiling design for the noninvasive prenatal diagnosis of congenital adrenal hyperplasia inherited with an autosomal recessive pattern (11).

In the present study, we attempted to demonstrate the feasibility of using the targeted massively parallel sequencing platform for carrier detection and noninvasive prenatal diagnosis of DMD.

MATERIAL AND METHODS

PATIENTS

The four DMD families receiving a prenatal diagnosis were prospectively recruited. Each family cohort consisted of a proband and the carrier mother. *DMD* mutations in the families included both large deletion/duplication and point mutations (Table 1). The experiment was designed and performed in two parts. First, genomic DNA from four mother and DMD proband pairs were sequenced to test whether an inherited *DMD* mutation from the carrier mother could be detected confidently. Second, maternal plasma DNA from the four carrier mothers at varying weeks of gestation was sequenced to determine whether inheritance of a *DMD* mutation from a carrier mother could be predicted in her fetus. Fetal genomic DNA obtained from either, chorionic villi sampling or amniocentesis was used to validate the results of the maternal plasma DNA sequencing. All procedures were performed as routine prenatal diagnosis. Additional informed consent was obtained for the study that used maternal plasma DNA and fetal DNA. The institutional review board approved the study protocol (IRB No. 1302-055-464).

TARGET ENRICHMENT AND MASSIVELY PARALLEL SEQUENCING

Maternal plasma (8–10 mL) was collected and prepared using the standard method as described previously (12). To construct the DNA library, we used the SureSelectXT Reagent Kit (Agilent Technologies, Santa Clara, CA) and 0.5–1 μg of plasma DNA for each case. Because the library-preparation section in the SureSelect protocol is designed primarily for genomic DNA, we modified it by diluting all reagents in the kit to prepare the plasma DNA library. This protocol is better suited for small amounts of input DNA. The adapter-ligated DNA was purified directly with the spin columns provided in the QIAquick PCR Purification

Kit (Qiagen, Hilden, Germany) without further size selection. Four-cycle PCR and SureSelect primers were then used to amplify the adapter-ligated DNA.

We quantified the DNA libraries using a Qubit 2.0 fluorometer (Invitrogen, Carlsbad, CA) and used the DNA 1000 Kit with a 2100 Bioanalyzer (Agilent) to check the size distribution of the libraries. We generated 0.3–0.5 μg of an amplified plasma DNA library for each sample, with an approximate mean size of 270 base pairs. Targeted sequence enrichment was performed using the SureSelect Custom Kit (Agilent). The custom capture probes targeted entire transcribed *DMD, ZFX*, and *ZFY* regions according to four gene databases (RefSeq, Ensembl, CCDS, and GENCODE) and were designed using Agilent SureDesign (https://earray.chem.agilent.com/suredesign). The following parameters were used for the capture design sequences: density, 2×; masking, least stringent; and boosting, balanced. We incubated 300 ng of the amplified plasma DNA library with the capture probes for 24 hours at 65 ℃, in accordance with the manufacturer's instructions. After hybridization, we selected the captured targets by pulling down the biotinylated probe–target hybrids with streptavidin-coated magnetic beads (Dynabeads M-280 Streptavidin; Invitrogen) and purified the targets using a MinElute PCR Purification Kit (Qiagen). Finally, we enriched the targeted DNA libraries using 12-cycle PCR amplification with SureSelect PCR primers (Agilent). The PCR products were purified using the QIAquick PCR Purification Kit. The library was paired-end sequenced on the Illumina HiSeq 2000 sequencing system. The sequenced paired-end reads were submitted to EBI European Nucleotide Archive (ENA) database with accession number PRJEB7629 (Direct access: http://www.ebi.ac.uk/ena/data/view/PRJEB7629)

VARIANT CALLING

Paired-end sequencing reads were aligned to the human genome (Genome Reference

Consortium Human Reference 37) with Bowtie2 aligner (v.2.2.3) (13). Picard Tools (http://picard.sourceforge.net) was used to remove PCR-duplicated reads, and duplicate-free BAM files were indexed by using SAMtools (v.0.1.19) (14). Local realignment around small insertions and deletions (indels), and base quality score recalibration were achieved using the Genome Analysis Toolkit (GATK, v.3.2-2). Variant calling was performed using GATK HaplotypeCaller. We filtered out low-quality variant calls using GATK VariantFiltration with parameters described by GATK Best Practice (http://www.broadinstitute.org/gatk/guide/best-practices) (154). Using our in-house script, we also filtered out variants with a genotype quality ≤30 and read depth ≤200. Lastly, we used ANNOVAR to annotate the unfiltered variants against the RefSeq gene set (16).

STRUCTURAL VARIATION DETECTION

Pindel (0.2.4.w) was used to detect structural variations (17). Only structural variations with a supportive read count ≥50, minimum length ≥1000 on the *DMD* gene were selected as pathogenic candidates. Compared with coverage plots visualized by the UCSC genome browser, large deletions/duplications were confirmed (18).

HAPLOTYPE CONSTRUCTION

Because of hemizygosity in males, we directly phased the maternal haplotypes of the *DMD* region. Using heterozygous single-nucleotide variants (SNVs) in the genomic DNA sequencing from the carrier mothers and probands, we classified the inherited haplotype that contained a deleterious mutation as haplotype A (HapA) and the other haplotype without a mutation as haplotype B (HapB).

MEASUREMENT OF FRACTIONAL FETAL DNA CONCENTRATION

In addition to the *DMD* gene, capturing the *ZFX* and *ZFY* genes provided a measurement of

the fractional fetal DNA concentration. Using mean read depth of two zinc finger genes ($\overline{\text{ZFX}}$ and $\overline{\text{ZFY}}$) with a minimum mapping quality score of 20 and base quality score of 20, we calculated the fractional fetal DNA concentration as:

$$\text{Fractional fetal DNA concentration} = \frac{2 \times \overline{\text{ZFY}}}{\overline{\text{ZFX}} + \overline{\text{ZFY}}} \times 100\%.$$

FETAL GENOTYPE PREDICTION

The *DMD* gene is known to have a high recombination rate, and tests to detect the recombination event and recombination point were performed before fetal genotype prediction. To prevent the occurrence of a prediction error for the recombination point because of outlier values originating from duplicated or repetitive regions, we used the R package 'qcc' for outlier removal (19). After outlier detection, we predicted the change point in the read fraction values using the R package 'bcp' (Bayesian change point) (20). After detection of the recombination event, we reconstructed haplotypes with and without deleterious *DMD* mutations and designated these as HapA* and HapB*, respectively. Subsequently, we predicted the fetal genotype by identifying the allele fraction imbalance between two haplotypes obtained from maternal plasma sequencing. Because the inherited allele would be overrepresented in relation to the fetal DNA fraction in the maternal plasma, the fetal genotype was determined by estimating which haplotype was overrepresented. If the overrepresented haplotype was the one harboring the *DMD* mutation, the fetus could be predicted to have inherited the *DMD* mutation (Fig. 1).

The statistical significance of the allele fraction imbalance was estimated using a one-tailed Student's paired *t* test or Wilcoxon signed-rank test depending on the assumption of normality. All statistical tests were performed with outlier-removed datasets.

RESULTS

CARRIER DETECTION FROM GENOMIC DNA SEQUENCING OF MOTHER AND PROBAND PAIRS

Targeted deep sequencing of four mother and proband pairs revealed uniform coverage across the *DMD* gene. A summary of the basic sequencing of the four pairs is provided (Table 2.). Large deletions/duplications were identified on visual inspection from a coverage plot across the *DMD* gene in both probands and carrier mothers (Fig. 2). The breakpoints were estimated successfully using the structural variation detection software, Pindel. Using visual inspection and breakpoint estimation, the predicted deleted or duplicated exons in all pairs were identical to the previous results detected using the multiple ligation-dependent probe amplification method. In agreement with the previous result, probands with a deletion mutation had nearly zero read depth at the deletion site, whereas the carriers with a deletion mutation had about half the read depth compared with the baseline read depth outside the deleted region (DMD-01 and DMD-03). The read depth height of the carrier with a duplication mutation was positioned between that of the baseline and proband with a duplication mutation (DMD-02). An inherited splice site mutation was also identified in the DMD-04-proband and the carrier mother (Fig. 3). The number of heterozygous or hemizygous SNVs in carriers and probands ranged from 700 to 1200 (Table 3). We successfully constructed two maternal haplotypes in the *DMD* gene using heterozygous SNVs and their proband haplotype. The mean read depth ratio of *ZFY* to *ZFX* ranged from 0.95 to 0.98 in male probands and was 0 in female carriers, suggesting that these zinc finger genes could be used as a reliable indicator of the fractional cffDNA concentration in the subsequent study using maternal plasma (Table 4).

FETAL GENOTYPE PREDICTION BY MATERNAL PLASMA DNA SEQUENCING

Seven plasma DNA samples obtained from four pregnant carriers at different gestational weeks were sequenced. The sequencing results showed uniform and high coverage across the *DMD* gene (Fig. 4). The number of SNVs and indels were also compatible with the genomic DNA sequencing data (Table 5). Fractional cffDNA concentration estimated by calculating the mean read depth of *ZFX* and *ZFY* ranged from 5.8% to 9.7%. Higher duplicated read rates (18–28%) were noted in seven plasma samples used for DNA sequencing (Table 6). This difference may have originated from additional PCR cycles used during library preparation and target enrichment because of the low concentration of the input plasma DNA.

Before examining the haplotype imbalance between the two phased maternal haplotypes, the recombination event within the *DMD* region was investigated using the R package, as described in Materials and Methods. The bcp algorithm estimated a significant change point in the read fraction in the sequencing data from DMD-02-9wk and DMD-02-12wk which were also evident from a scatterplot of the read fraction distribution of the phased haplotype (Fig. 5). Subsequent analysis suggested the recombination point between chromosomal X positions 32,321,115 and 32,346,373 based on the bcp algorithm. The haplotypes of DMD-02-9wk and DMD-02-12wk sequencing data were reconstructed using the recombination point information. Because all the other plasma sequencing data revealed one large segment of two haplotypes, all remaining families, with the exception of DMD-02, used the same haplotype that was phased from the proband sequencing data (Fig. 6A–C).We next attempted to predict the fetal genotype by comparing the allele fraction between the two haplotypes in the maternal plasma. In both DMD-01-6wk and DMD-01-17wk, the allele fraction of HapB was significantly higher, indicating inheritance of the non-mutated haplotype by the fetus (Fig. 7A). All remaining samples supported the inheritance of a mutated haplotype by the fetuses, including DMD-02-9wk and DMD-02-12wk, the haplotype of which were reconstructed based on a recombination event prediction (Fig. 7B-D). All the allele fraction

differences and statistical test results are provided (Table 6). The fetal genotypes predicted from the four DMD families matched exactly the fetal genomic DNA sequencing data (Fig. 8).

DISCUSSION

In the present study, we have demonstrated that targeted deep sequencing makes feasible not only genetic diagnosis of a DMD patient but also carrier detection and noninvasive prenatal diagnosis. Considering that the prenatal diagnosis of monogenic disorders is still performed in a family-based setting in which a genetically confirmed proband or carrier has been identified, this method has a practical advantage that proband diagnosis, carrier detection, and noninvasive prenatal diagnosis can be accomplished efficiently with a single platform.

For clinical implementation of noninvasive prenatal diagnosis of X-linked recessive disorders including DMD, refining the detection method of dosage imbalance caused by presence of a small fraction of cffDNA is a critical step in identifying the maternally derived genotype. Tsui et al. used digital PCR to detect the slight overrepresentation of a *F8* mutation in pregnant women who are carriers of hemophilia mutations (12). The relative mutation dosage analysis based on a sequential probability ratio test was used. Although it is simple and does not require haplotype information, the detecting probe must be individualized according to the specific mutation type. Further, multiple tests of each sample should be conducted to obtain sufficient statistical power, especially when the fetal DNA fraction is low, as in the early gestational weeks. Relative haplotype dosage analysis may be an alternative option for identifying the slight overrepresentation of an inherited maternal mutation or allele because the genome-wide or targeted massively parallel sequencing approach can produce many informative SNVs for haplotyping. Lam et al. used this approach to identify the maternal inheritance of a mutation in a β-thalassemia model (9). Although β-thalassemia is a disease with autosomal recessive inheritance, a method that identifies the maternal inheritance pattern would be equally applicable to X-linked recessive diseases. Instead of using the separate haplotyping method employed by Lam et al., New et al. sequenced a parent and patient trio, and then used the resulting haplotype information for maternal DNA analysis in the targeted

platform used in the trio analysis (11). This diagnostic flow could be incorporated more easily into the current genetic diagnosis and counseling process.

Several gene-specific factors inherent to *DMD* should be considered for clinical implementation of noninvasive prenatal diagnosis. First, large deletion/duplication mutations constitute about two-thirds of the *DMD* mutation spectrum. Because a dosage imbalance already exists in DMD carriers with a large deletion/duplication mutation, it would be difficult to perform relative mutation dosage analysis using digital PCR. Thus, measuring the haplotype imbalance of the *DMD* region outside the mutation would be plausible. Second, the recombination rate of *DMD* is about four times higher compared with chromosome X and with the whole genome: 4.73 cM/Mb, 1.21 cM/Mb, and 1.26 cM/Mb, respectively (21, 22). If a recombination event occurred within the *DMD* region, it would greatly affect the dosage imbalance analysis and could result in an incorrect prediction. Fetal genotype prediction without considering the inheritance of recombinant haplotypes would cause nonsignificant haplotype imbalance or contradictory result; it is possible to misdiagnose a non-DMD fetus as harboring a *DMD* mutation and vice versa. Thus, targeting the whole *DMD* region with a tiling design is preferable to ensure the reliable detection of a recombination event within the *DMD* region. Third, because the *DMD* region is hemizygous for male probands, all phased heterozygous SNVs are informative for dosage imbalance analysis provided that the recombination event is checked and corrected for before analysis.

Considering the abovementioned gene-specific factors, we postulate that our approach targeting the whole *DMD* region with tiling design provide the best approach for noninvasive prenatal diagnosis of DMD. Instead of using multiple haplotype blocks for repeated relative haplotype dosage analysis (5, 11), we hypothesized that the whole *DMD* gene could be analyzed as one large haplotype block, and we compared directly the allele fractions of two

phased maternal heterozygous alleles. Although the study reported by New et al., in which multiple haplotype blocks were used, demonstrated the clinical applicability of that design to an array of autosomal recessive disorders, we believe that the current approach using one large haplotype block is a simpler and more straightforward method, at least for DMD. Although the fetal genotypes from all four DMD families were accurately predicted, the application of the current approach to routine molecular testing may have several limitations, in particular regarding noninvasive prenatal diagnosis. First, the current proband-based phasing approach may lead to misdiagnosis if separate recombination occurs both in a proband and in a new fetus. In addition, it was basically impossible to discern whether the recombination event observed in the DMD-02 family had occurred in the proband or in the new fetus using the current proband-based phasing approach, although the inherited fetal genotype could still be correctly predicted. This disadvantage could be partly overcome by adding the grandfather to maternal haplotype phasing, as the maternal X haplotype inherited from the grandfather would theoretically be free of recombination. This approach of using the grandparents for phasing was introduced in a recent article by Meng et al. (10) Second, as female patients or carriers of DMD mutations with no known proband have been increasingly identified, the current proband-based phasing approach also has a practical limitation in that setting. An alternative method that uses other family members for maternal haplotype phasing could be introduced to overcome this limitation, although it may still not be applicable to all at-risk couples without a proband. A grandfather or normal male child may be used for phasing via the same approach. Moreover, a carrier female or normal female child may also be used for phasing when the paternal genotype is available. Third, the optimal timing of testing and the least fetal DNA fraction required should be determined and validated in extended DMD families with various DMD mutations. In the current study, the earliest gestational time and the lowest fetal fraction that allowed successful noninvasive prenatal

diagnosis were 6 weeks and 5 days and 5.8%, respectively. New et al. reported a successful case of prenatal diagnosis at 5 weeks and 6 days of gestation with a fetal fraction of 1.4% (11). We believe that our results are compatible with those of the study performed by New et al. in terms of resolution, considering that 5 to 6 weeks of gestation might be the earliest period at which prenatal tests may be offered. However, data collection in extended families is needed, as the current study used only four families compared with the 14 families reported by New et al. Fourth, because the informative SNVs that are required for dosage imbalance analysis of maternally inherited alleles might be limited in number and located at greater distance, the current approach using one large haplotype block may be biased by false-recombination prediction. Thus, the extension of the applicability of the current approach to autosomal recessive disorders should be demonstrated separately.

Besides the technical considerations, there are complex ethical and socioeconomic issues to be addressed before such an approach can be implemented in the clinic. Currently, no curative therapy is available for DMD, although some therapeutic molecules are under clinical trial (23). However, considering the current method is best fitted to clinical circumstances in which the presence of an affected proband is the reason for prenatal testing, noninvasive determination of fetal genotype in the early gestational weeks could provide an autonomy-based reproductive option to the parents.

Despite the need to overcome these various hurdles, our approach for the comprehensive genetic diagnosis of the proband and noninvasive prenatal diagnosis using a single massively parallel targeted sequencing platform may provide a practical model for implementation of next-generation sequencing technology to clinical genetics.

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, Lo YM. Clinical applications of maternal plasma fetal DNA analysis: Translating the fruits of 15 years of research. Clin Chem Lab Med 2013;51:197-204.
3. Lo YM. Non-invasive prenatal testing using massively parallel sequencing of maternal plasma DNA: From molecular karyotyping to fetal whole-genome sequencing. Reprod Biomed Online 2013;27:593-8.
4. American College of O, Gynecologists Committee on G. Committee opinion no. 545: Noninvasive prenatal testing for fetal aneuploidy. Obstet Gynecol 2012;120:1532-4.
5. 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.
6. Liao GJ, Lun FM, Zheng YW, Chan KC, Leung TY, Lau TK, et al. Targeted massively parallel sequencing of maternal plasma DNA permits efficient and unbiased detection of fetal alleles. Clin Chem 2011;57:92-101.
7. Kitzman JO, Snyder MW, Ventura M, Lewis AP, Qiu R, Simmons LE, et al. Noninvasive whole-genome sequencing of a human fetus. Sci Transl Med 2012;4:137ra76.
8. Lim BC, Lee S, Shin JY, Kim JI, Hwang H, Kim KJ, et al. Genetic diagnosis of duchenne and becker muscular dystrophy using next-generation sequencing technology: Comprehensive mutational search in a single platform. J Med Genet 2011;48:731-6.
9. 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.

10. Meng M, Li X, Ge H, Chen F, Han M, Zhang Y, et al. Noninvasive prenatal testing for autosomal recessive conditions by maternal plasma sequencing in a case of congenital deafness. Genet Med 2014;99:E1022-30.
11. 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.
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. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods 2012;9:357-9.
14. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and samtools. Bioinformatics 2009;25:2078-9.
15. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491-8.
16. Wang K, Li M, Hakonarson H. Annovar: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
17. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: A pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 2009;25:2865-71.
18. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at ucsc. Genome Res 2002;12:996-1006.
19. Scrucca, L. qcc: an R package for quality control charting and statistical process

control. . R News 2004;4/1:11-7.

20. Erdman C, Emerson JW. A fast bayesian change point analysis for the segmentation of microarray data. Bioinformatics 2008;24:2143-8.

21. Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen CF, et al. Comparative recombination rates in the rat, mouse, and human genomes. Genome Res 2004;14:528-38.

22. International HapMap C. The international hapmap project. Nature 2003;426:789-96.

23. Jarmin S, Kymalainen H, Popplewell L, Dickson G. New developments in the use of gene therapy to treat duchenne muscular dystrophy. Expert Opin Biol Ther 2014;14:209-30.

24. Tengfei Yin, Dianne Cook, Michael Lawrence. "ggbio: an R package for extending the grammar of graphics for genomic data." Genome Biol 2012 13: R77

25. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative genomics viewer (igv): High-performance genomics data visualization and exploration. Brief Bioinform 2013;14:178-92.

Table 1. *DMD* mutation status of the study cohort and types of samples sequenced.

Study number	*DMD* mutation (By MLPA[a] or Sanger sequencing)		Genomic DNA sequencing		Maternal plasma DNA sequencing(gestational age)	Fetal DNA sequencing
	Proband[b]	Mother	Proband	Mother		
DMD-01	exon 49–52 deletion	Carrier	+	+	+ (6 wk 5d, 17 wk 1d)	+
DMD-02	exon 2 duplication	Carrier	+	+	+ (9 wk 3d, 12 wk 1d)	+
DMD-03	exon 3-7 deletion	Carrier	+	+	+ (8 wk 5d, 11 wk 3d)	+
DMD-04	c.649+2T>C	Carrier	+	+	+ (7 wk 1d)	+

[a] MLPA, multiple ligation-dependent probe amplification.

[b] All probands were male.

Table 2. Targeted massively parallel sequencing summary of genomic and maternal plasma DNA sequencing.

Sample	Total reads	Total reads mapped to HG19	Total reads mapped to HG19 (%)	Total reads mapped to Target	Total reads mapped to Target (%)	Covered bait bases >= 30 (%)	Mean read depth (x)
DMD-01-mother	67913844	67040224	98.71%	67913844	38.80	97.7	1211.23
DMD-01-proband	68111649	67032588	98.42%	68111649	25.45	94.9	810.17
DMD-01-fetus	62140714	61254461	98.57%	62140714	23.91	99.7	693.84
DMD-02-mother	63394776	62700321	98.90%	63394776	41.04	97.7	1196.67
DMD-02-proband	69839872	68993794	98.79%	69839872	26.96	99.6	880.73
DMD-02-fetus	53482735	52826547	98.77%	53482735	21.04	99.7	527.45
DMD-03-mother	66536460	65974792	99.16%	66536460	45.49	97.6	1316.77
DMD-03-proband	30627950	30326923	99.02%	30627950	28.40	92.1	383.68
DMD-03-fetus	33642168	33334631	99.09%	33642168	23.18	92.2	338.13
DMD-04-mother	34191912	33941764	99.27%	34191912	49.51	97.4	744.87
DMD-04-proband	32498580	32193877	99.06%	32498580	33.96	99	486.78
DMD-04-fetus	29799572	29478604	98.92%	29799572	26.45	99.2	249.23
DMD-01-6wk	54828936	53912045	98.33%	54828936	18.41	97.7	465.61
DMD-01-17wk	53776131	52978358	98.52%	53776131	21.30	98	529.38

DMD-02-9wk	50420886	49643304	98.46%	50420886	18.91	98	440.4
DMD-02-12wk	65047893	64180117	98.67%	65047893	26.40	98.6	792.31
DMD-03-8wk	68397568	67646003	98.90%	68397568	37.79	98.3	698.06
DMD-03-11wk	68874610	68144635	98.94%	68874610	37.89	98.9	724.04
DMD-04-7wk	65317530	64610871	98.92%	65317530	38.70	98.1	687.24

[a] Genomic DNA samples of DMD-01-fetus and DMD-02-fetus were collected by chorionic villus sampling and amniocentesis, respectively.

Table 3. Variant count summary of the samples in the present study.

Sample	Total variants	SNVs	Heterozygous SNVs	Homozygouns SNVs	Indels
DMD-01-mother	1459	1402	766	636	57
DMD-01-proband	1066	1026	–	1026	40
DMD-01-fetus	1057	1018	–	1018	39
DMD-02-mother	1681	1631	1023	608	50
DMD-02-proband	1209	1166	–	1166	43
DMD-02-fetus	1192	1149	–	1148	43
DMD-03-mother	1376	1326	724	602	50
DMD-03-proband	1024	994	–	994	30
DMD-03-fetus	1011	982	–	982	29
DMD-04-mother	1656	1581	900	681	75
DMD-04-proband	1106	1057	–	1057	49
DMD-04-fetus	952	920	–	920	32
DMD-01-6wk	1476	1415	781	634	61
DMD-01-17wk	1482	1423	783	640	59
DMD-02-9wk	1722	1662	1042	620	60
DMD-02-12wk	1732	1678	1052	626	54
DMD-03-8wk	1460	1408	761	647	52
DMD-03-11wk	1456	1405	760	645	51
DMD-04-7wk	1686	1611	914	697	75

Table 4. Mean read depth of two zinc finger genes and ZFX to ZFY ratio of the study samples.

Sample	Mean read depth		*ZFY* to *ZFX* ratio	Fractional fetal DNA concentration (%)
	ZFX	*ZFY*		
DMD-01-mother	1150.06	0.92	0	–
DMD-01-proband	779.27	759.14	0.97	–
DMD-01-fetus[a]	637.27	621.54	0.98	–
DMD-02-mother	1105.32	0.91	0	–
DMD-02-pronband	794.58	770.91	0.97	–
DMD-02-fetus[a]	504.52	491.97	0.98	–
DMD-03-mother	1293.16	1.2	0	–
DMD-03-proband	389.81	375.99	0.96	–
DMD-03-fetus	333.72	322.74	0.96	–
DMD-04-mother	649.85	0.75	0	–
DMD-04-proband	454.98	435.24	0.95	–
DMD-04-fetus	228.71	220.37	0.96	–
DMD-01-6wk	442.57	12.88	0.04	5.6
DMD-01-17wk	502.61	20.25	0.02	7.7
DMD-02-9wk	412.21	20.01	0.04	9.2
DMD-02-12wk	751.59	26.66	0.03	6.8
DMD-03-8wk	677.24	22.17	0.03	6.3
DMD-03-11wk	708.2	32.62	0.04	8.8
DMD-04-7wk	634.8	20.45	0.03	6.2

Table 5. Duplicated read rate of the samples in the present study.

Sample	Total reads	Total reads with Duplicated reads	Duplicated reads	Duplicated reads (%)
DMD-01-mother	67,913,739	73,055,256	5,141,517	7.03
DMD-01-proband	68,111,501	72,365,082	4,253,581	5.87
DMD-01-fetus	62,140,727	69,473,676	7,332,949	10.55
DMD-02-mother	63,394,802	67,051,900	3,657,098	5.45
DMD-02-proband	69,840,025	74,232,768	4,392,743	5.91
DMD-02-fetus	53,482,707	59,063,358	5,580,651	9.44
DMD-03-mother	65027479	66536460	1508981	2.26
DMD-03-proband	30245725	30627950	382225	1.24
DMD-03-fetus	32803261	33642168	838907	2.49
DMD-04-mother	33577754	34191912	614158	1.79
DMD-04-proband	32036957	32498580	461623	1.42
DMD-04-fetus	26299406	29799572	3500166	11.74
DMD-01-6wk	54,828,497	74,741,546	19,913,049	26.64
DMD-01-17wk	53,776,065	72,911,708	19,135,643	26.24
DMD-02-9wk	50,421,216	69,319,082	18,897,866	27.26
DMD-02-12wk	65,047,677	80,532,264	15,484,587	19.22
DMD-03-8wk	55738457	68397568	12659111	18.5
DMD-03-11wk	56245002	68874610	12629608	18.33
DMD-04-7wk	53312149	65317530	12005381	18.38

Table 6. Allele fraction of two phased haplotypes in genomic and maternal plasma DNA sequencing data.

Sample	Allele fraction		Allele fraction difference (%)	Fractional fetal DNA concentration (%)	Paired-difference tests (*P*-value)	
	Haplotype A	Haplotype B			Student's paired *t* test	Wilcoxon signed-rank test
DMD-01-6wk	0.475	0.525	5 .0	5.8	<2.2e–16	< 2.2e–16
DMD-01-17wk	0.471	0.529	5.6	8	<2.2e–16	< 2.2e–16
DMD-02-9wk[a]	0.522	0.477	4.4	9.73	<2.2e–16	< 2.2e–16
DMD-02-12wk[a]	0.514	0.485	2.9	7.1	<2.2e–16	< 2.2e–16
DMD-03-8wk	0.528	0.472	5.6	6.3	<2.2e–16	< 2.2e–16
DMD-03-11wk	0.525	0.475	4.9	8.8	<2.2e–16	< 2.2e–16
DMD-04-7wk	0.514	0.485	2.9	6.2	<2.2e–16	< 2.2e–16

Fig. 1. Illustration of the workflow of a comprehensive genetic diagnosis in a DMD carrier, proband, and fetus.

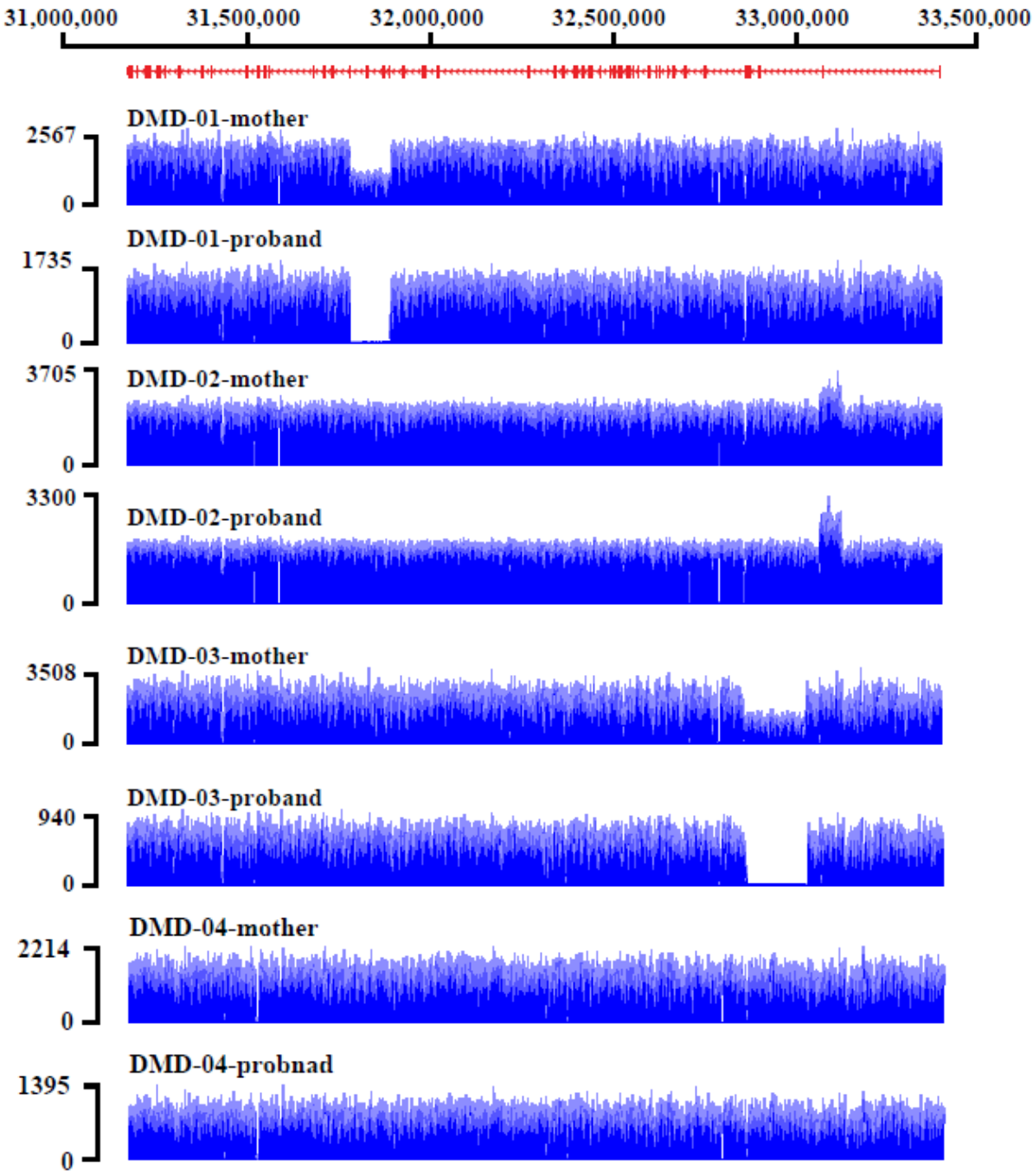


Fig. 2. Coverage plots of genomic DNA sequencing data.

Coverage plots of genomic DNA samples. Each position corresponding to a large deletion/duplication was completely matched in pairs. The red vertical bars represent the 79 exons of the DMD gene.

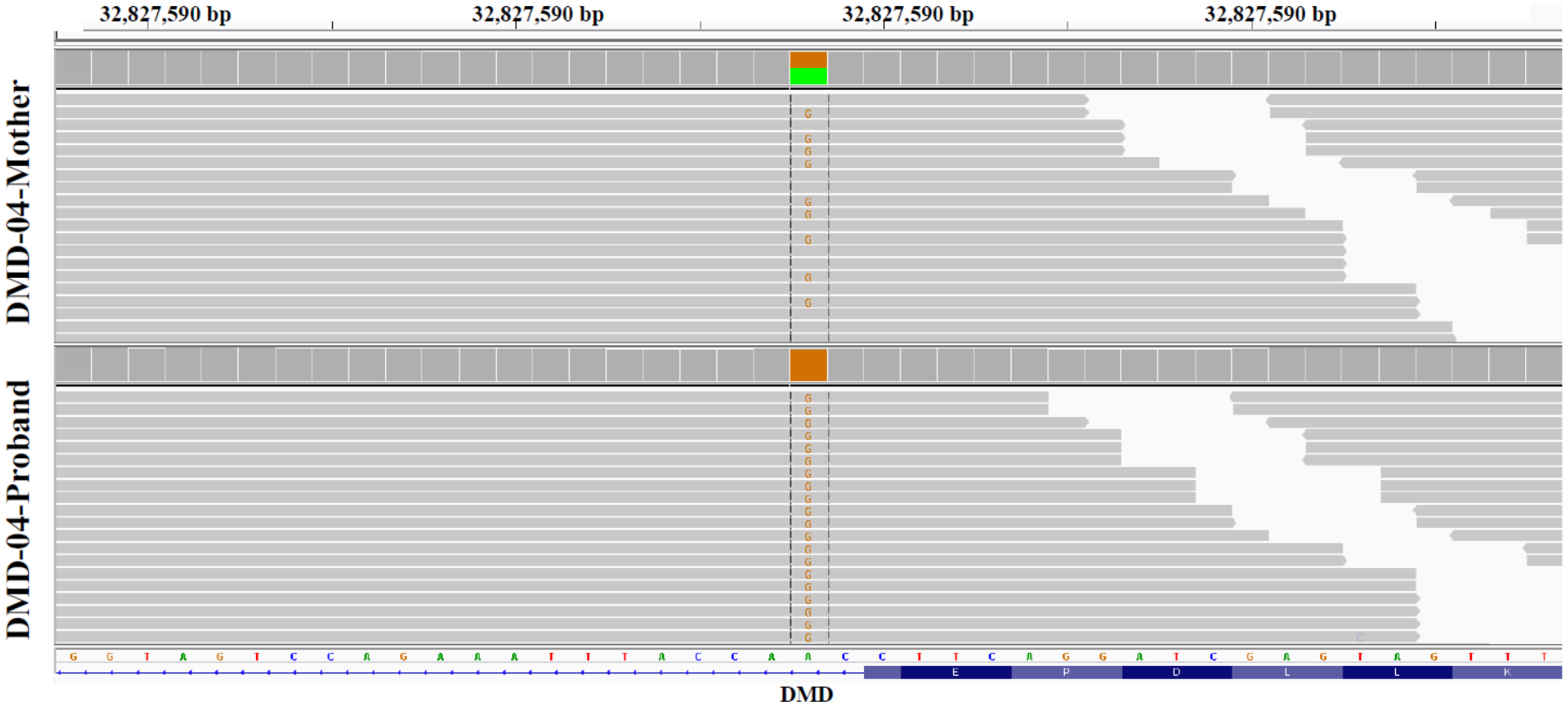


Fig. 3. Identification of single-nucleotide variants in DMD-04 pair.

Targeted deep-sequencing result visualized by the Integrative Genomics Viewer (25).

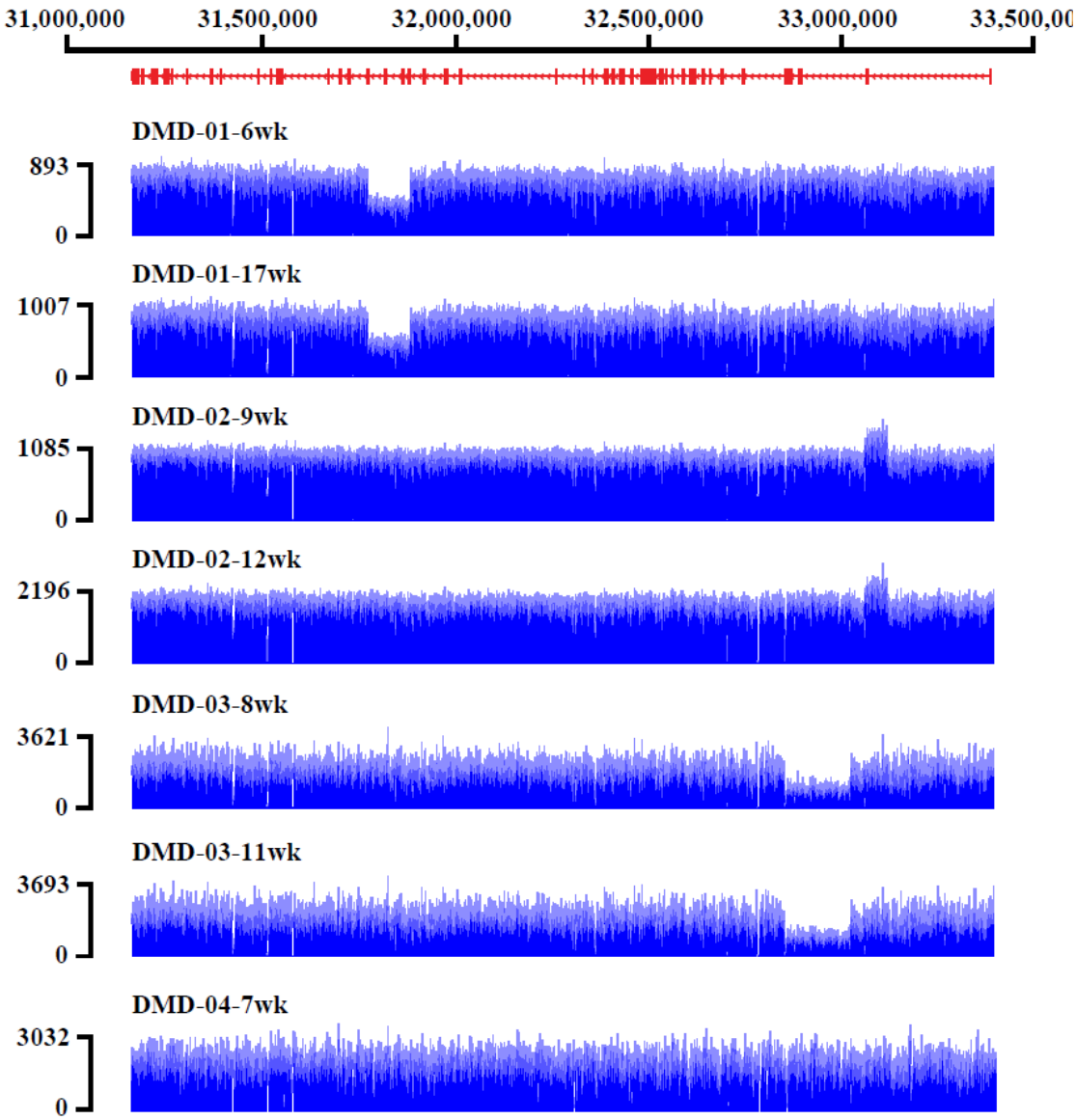


Fig. 4. Coverage plots of plasma DNA sequencing data.

Coverage plots of plasma DNA samples. Four pregnant carriers at different gestational ages exhibited uniform coverage and identical deletions/duplications detected in genomic DNA samples. The red vertical bars represent the 79 exons of the DMD gene**.**

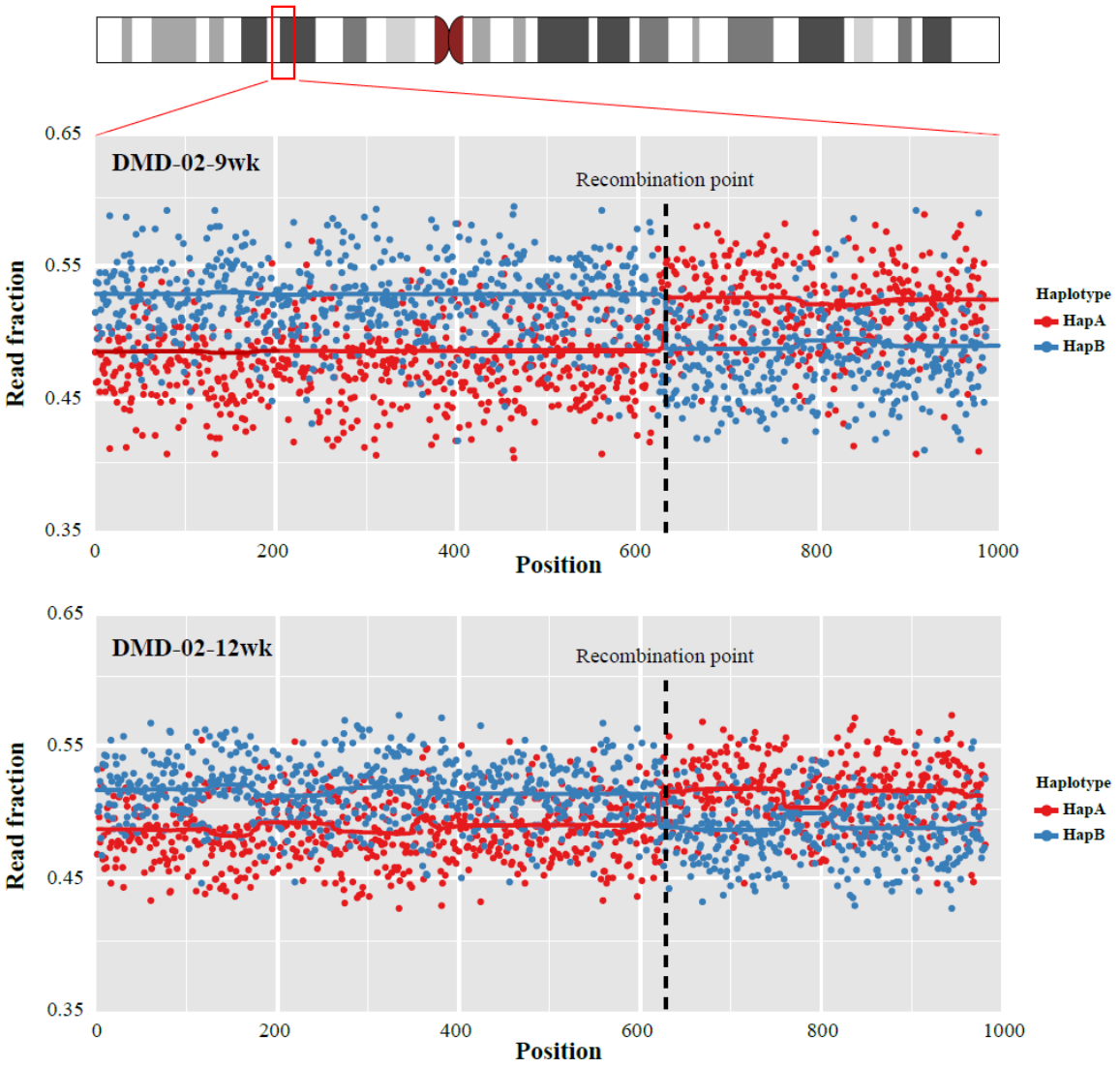


Fig. 5. Detection of a recombination event and haplotype reconstruction in the DMD-02 family.

Read fraction distribution of two haplotypes. Each haplotype was divided into two large segments. The black dotted lines indicate the recombination point predicted by the bcp algorithm. The karyogram representing chromosome X was generated by ggbio (24).

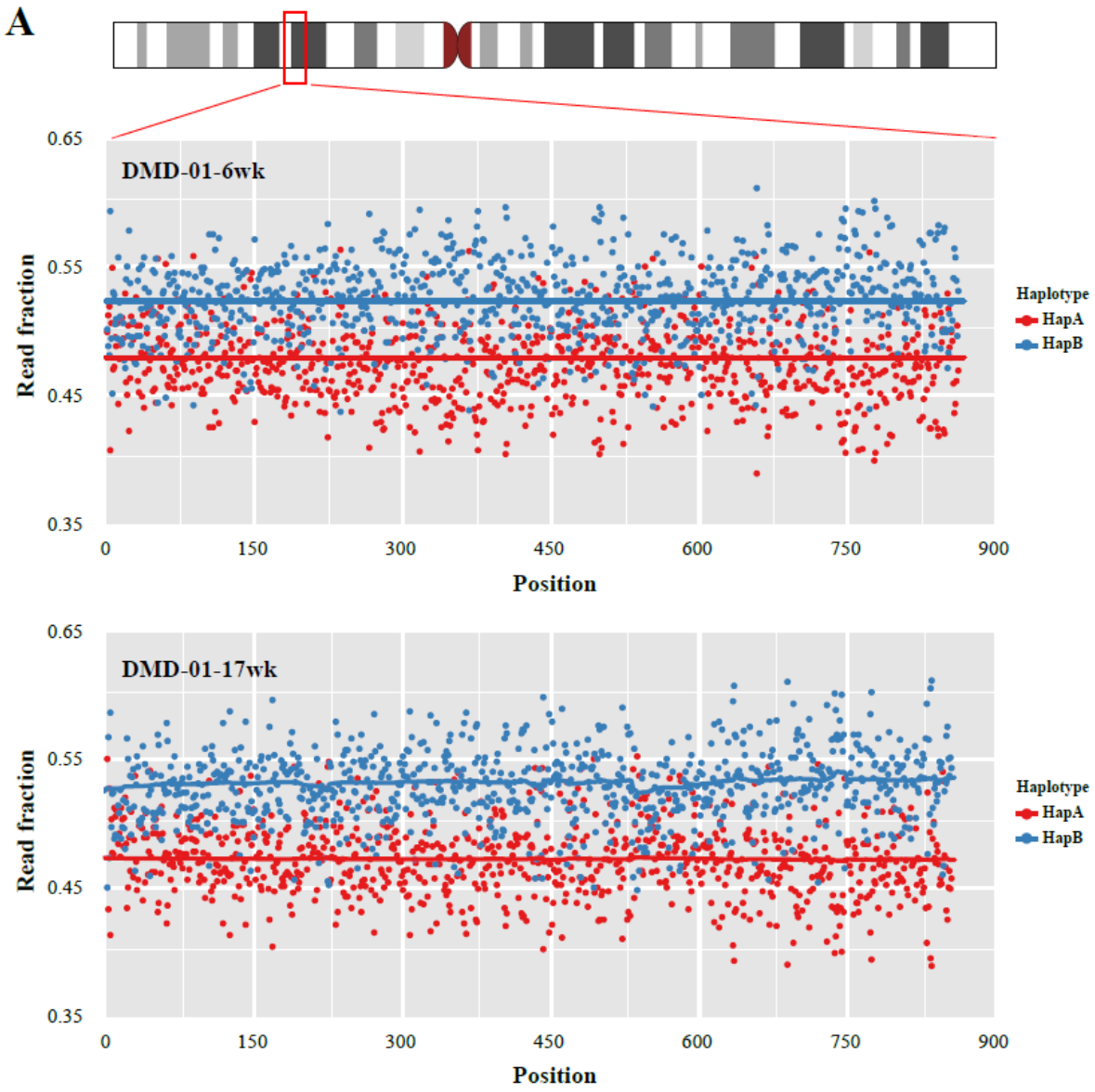
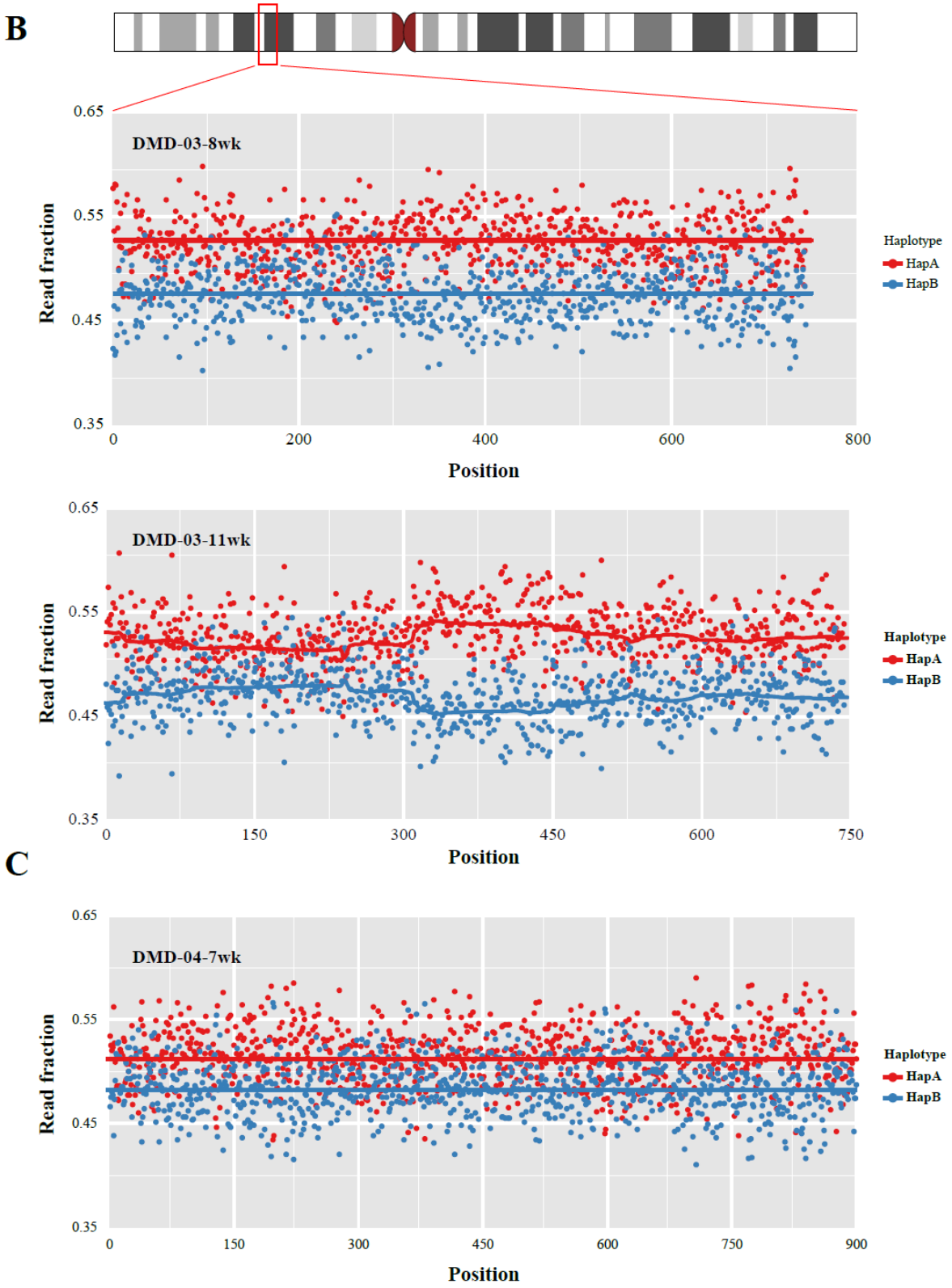


Fig. 6. Read fraction distribution of plasma samples without recombination events.

The distribution of read fractions represents one large segment of each haplotype. (A) DMD-01 family, (B) DMD-03 family, and (C) DMD-04 family.


B
DMD-03-8wk
Read fraction
0.65
0.55
0.45
0.35
0
200
400
600
800
Position
Haplotype
HapA
HapB
DMD-03-11wk
Read fraction
0.65
0.55
0.45
0.35
0
150
300
450
600
750
Position
Haplotype
HapA
HapB
C
DMD-04-7wk
Read fraction
0.65
0.55
0.45
0.35
0
150
300
450
600
750
900
Position
Haplotype
HapA
HapB

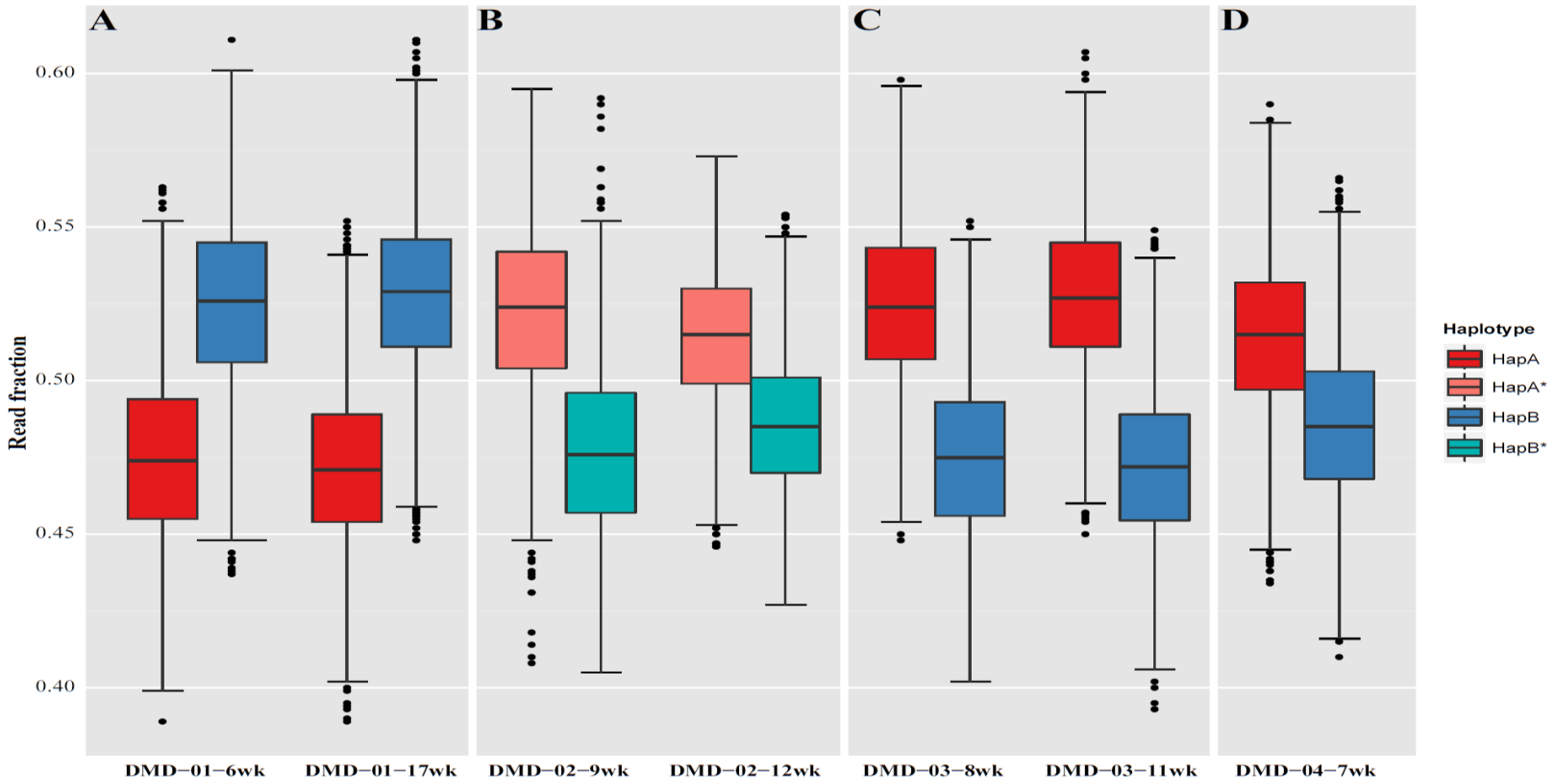


Fig. 7. Fetal genotype prediction.

(A) HapB is overrepresented in DMD-01 maternal plasma samples (allele fraction differences: 5.0% and 5.6%). HapA or HapA* are overrepresented in the remaining maternal plasma samples. (B) DMD-02 (allele fraction differences: 4.4% and 2.9%). (C) DMD-03 (allele fraction differences: 5.6% and 4.9%). (D) DMD-04 (allele fraction difference: 2.9%).

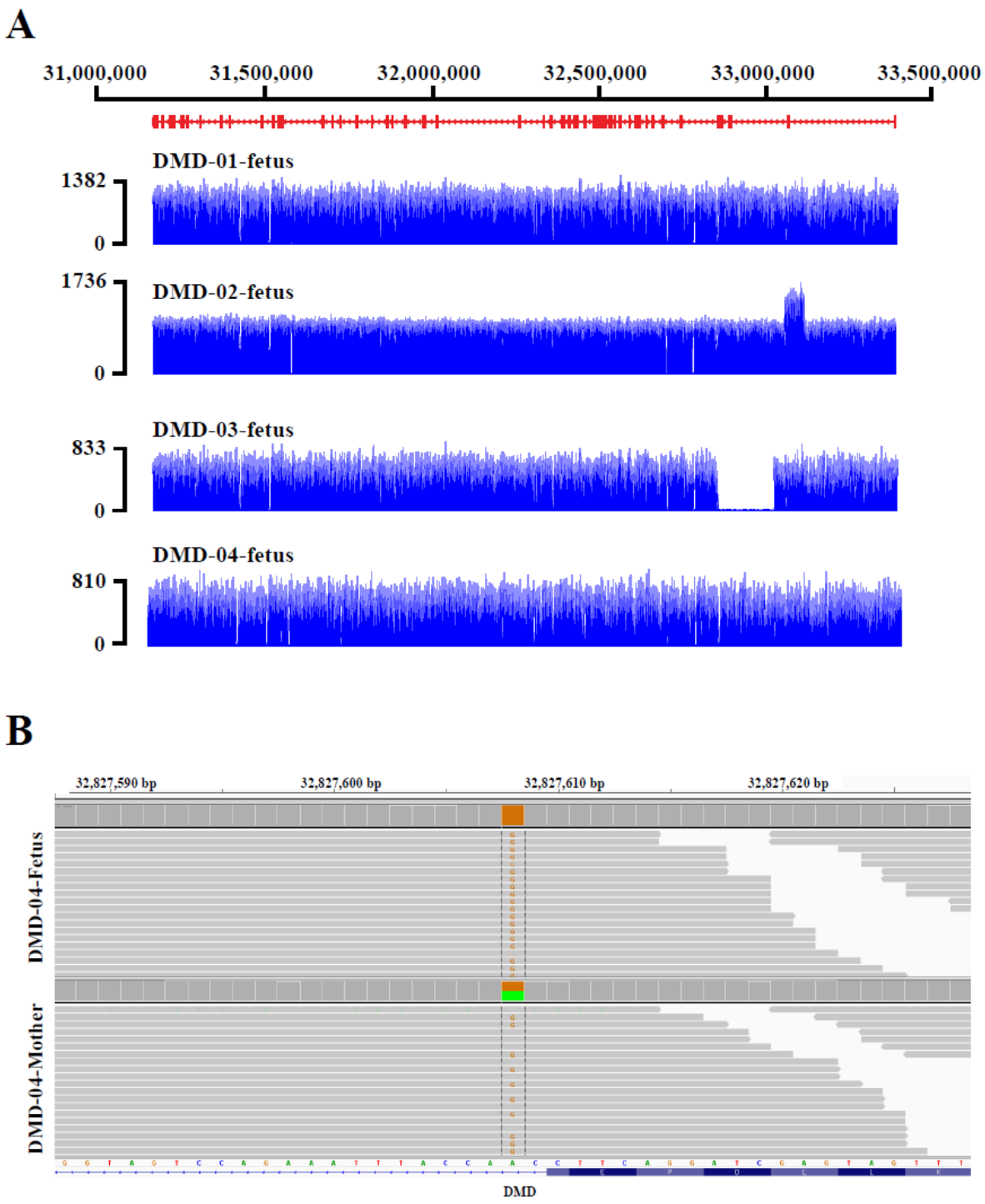


Fig. 8. Fetal genotype validation.

Fetal genotypes predicted using plasma DNA samples were successfully validated by fetal genomic DNA obtained via either chorionic villi sampling or amniocentesis. (A) Large deletion/duplication. (B) Splice-site mutation.

한글 초록

배경: 산모혈장에 존재하는 태아 DNA와 차세대시퀀싱 기술을 결합하여 비침습적 산전진단 방법을 개발하려는 연구가 활발히 진행되고 있다. 사전연구에서 차세대시퀀싱을 이용한 Duchenne형 근이영양증의 유전진단 플랫폼을 개발한 바 있는데, 이를 토대로 보인자 확인 및 비침습적 산전진단법 개발까지 확대 응용하는 방법을 개발하고자 하였다.

방법: 디스트로핀 유전자 전장을 포함하도록 타겟 지역을 설계하여 캡쳐 키트를 제작하였다. 이미알려진 4명의 보인자 산모와 환자에서 보인자 상태의 진단이 가능한지 차세대시퀀싱을 통해서 분석하였다. 상기 4가계에서 다양한 임신주수 에서 산모 혈장을 수집하여 차세대시퀀싱 및 정해진 알고리듬에 따라 태아의 유전형 예측이 가능한지 확인한다. 태아의 유전형은 산모의 두 하프로타입의 양을 수치화 하여 비교하여 산모혈장 시퀀싱 데이터에서 양이 많은 하프로타입이 태아에 전달된 것으로 예측한다. 하프로타입의 구분은 환자의 시퀀싱 데이터을 이용하고, 이 때 유전자 재조합의 가능성을 고려하여 분석한다.

결과: 상기 플랫폼을 이용하여 디스트로핀 유전자의 거대중복, 거대결실, 점돌연변이 모두에서 산모의 보인자 상태를 정확하게 확인할 수 있었다. 산모혈장 시퀀싱 데이터를 통해서도 4태아 모두에서 유전형을 정확하게 예측하였으며, 가장 일찍 예측 된 임신주수는 6주 5일 이었다. 한 가계에서는 유전자 재조합이 발생한 것이 확인 되었으며, 이를 토대로 교정한 후에 정확하게 예측이 가능하였다.

결론: 산모혈장의 태아 DNA 및 차세대시퀀싱 방법을 이용하여 Duchene형 근이영양증의 비침습적산전진단이 가능함을 증명하였다. 동일한 원리를 이용하여 X

연관 열성 유전양식의 기타 질환에도 확대적용을 기대할 수 있다.

주요어: 비침습적 산전진단, Duchenne형 근이영양증, 태아 혈장 DNA, 차세대시퀀싱

학번: 2009-30580