ARTICLE

Identification and characterization of two DMD pedigrees with large inversion mutations based on a long-read sequencing pipeline

Chang Geng[1,6], Ciliu Zhang[2,6], Pidong Li[3,6], Yuanren Tong[1], Baosheng Zhu[4], Jing He[4], Yanhuan Zhao[1], Fengxia Yao[5], Li-Ying Cui[1], Fan Liang[3], Yang Wang[3], Yaru Wang[3], Hongshuai Jin[3], Dandan Lang[3], Shanlin Liu[3], Depeng Wang[3], Min S. Park[3], Lin Chen[1], Jing Peng[2]✉ and Yi Dai[1]✉

Pathogenic large inversions are rarely reported on *DMD* gene due to the lack of effective detection methods. Here we report two DMD pedigrees and proposed a reliable pipeline to define large inversions in DMD patients. In the first pedigree, conventional approaches including multiplex ligation-dependent probe amplification, and whole-exome sequencing by next generation sequencing were failed to detect any pathologic variant. Then an advanced analysis pipeline which consists of RNA-seq, cDNA array capture sequencing, optical mapping, long-read sequencing was built. RNA-seq and cDNA capture sequencing showed a complete absence of transcripts of exons 3–55. Optical mapping identified a 55 Mb pericentric inversion between Xp21 and Xq21. Subsequently, long-read sequencing and Sanger sequencing determined the inversion breakpoints at 32,915,769 and 87,989,324 of the X chromosomes. In the second pedigree, long-read sequencing was directly conducted and Sanger sequencing was performed to verify the mutation. Long-read sequencing and Sanger sequencing found breakpoints at 32,581,576 and 127,797,236 on *DMD* gene directly. In conclusion, large inversion might be a rare but important mutation type in *DMD* gene. An effective pipeline was built in detecting large inversion mutations based on long-read sequencing platforms.

European Journal of Human Genetics (2023) 31:504–511; https://doi.org/10.1038/s41431-022-01190-y

INTRODUCTION

Duchenne muscular dystrophy (DMD) is an X-chromosome-linked muscle disease caused by mutations in *DMD* [1–4]. Many *DMD* mutations have been identified and been known to produce non- or partially functional dystrophins, which is essential to attach contractile microfilaments to sarcolemma [2, 4]. Whether mutations disrupt the reading frame determines the severity of the pathophysiology of DMD patients [5, 6]. The absence of this functional protein leads to continuous degeneration of muscle tissue, including skeletal muscle, myocardium and smooth muscle, and orthopedic and respiratory complications, thus causing high morbidity and mortality in DMD patients [2]. These patients are typically identified between 2 and 5 years of age and exhibit hypertrophy in calf and arm muscles, waddling gate, and insufficient agility in running or climbing stairs [2]. Some presymptomatic patients are also diagnosed earlier than 2 years of age due to known family history or high level of serum creatine kinase (CK) [2].

Over 7000 *DMD* mutations have been identified throughout 79 exons that span 2.5 Mb genomic DNA at Xp21 [7, 8]. Among these mutations, large deletions are the most common, accounting for 65% of mutations. The major deletion hotspot is located between exon 42 and 55 and 75% of deletions cluster in this region [9]. The rest of mutations are single nucleotide variants (SNVs), insertions of various sizes, duplications, and balanced translocations or inversions at very low frequency. Although the majority of these *DMD* mutations are easily identified by genetic diagnosis methods such as multiplex ligation-dependent probe amplification (MLPA) and whole exome sequencing (WES) based on next generation sequencing (NGS), rarely other mutations are detected by karyotyping [10], fluorescence in situ hybridization (FISH) [11], and chromosome array [12]. However, these conventional genetic tests sometimes fail to identify pathogenic mutations.

The recent development in advanced DNA sequencing and genome mapping technologies are revolutionizing the field of genome science and genomic medicine [13, 14]. The long-read sequencing platforms, such as single-molecule real-time sequencing (SMRT) [15] and nanopore sequencing [16] help us find many genetic and epigenetic mutations that have evaded detection by short-read sequencing platforms. These new technologies can identify complex mutations and structural variants (SVs) which were not previously found because of low sequencing depth or

[1]Department of Neurology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, 100730 Beijing, China. [2]Department of Pediatrics, Xiangya Hospital, Central South University, Changsha 410008 Hunan, China. [3]GrandOmics Biosciences, Beijing, China. [4]Department of Medical Genetics, The First People's Hospital of Yunnan Province, Kunming 650032 Yunnan, China. [5]Laboratory of Clinical Genetics, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, 100730 Beijing, China. [6]These authors contributed equally: Chang Geng, Ciliu Zhang, Pidong Li. ✉email: pengjing4346@163.com; pumchdy@pumch.cn

Received: 3 January 2022 Revised: 25 August 2022 Accepted: 5 September 2022
Published online: 5 October 2022

low sequence complexity around mutations. Recently, the optical mapping was proven effective in identifying complex variants in *DMD* [17]. Long-read DNA sequencing technology also aided the discovery of a large number of new SVs in the human genome [18–21].

Since advanced genetic tests described above were developed these years, we further built an advanced analysis pipeline which consists of RNA-seq, cDNA capture sequencing, optical mapping with Bionano platform, and nanopore-based long-read sequencing with ONT platform. Using the pipeline, we successfully identified a large inversion directly linked to the patient's dystrophinopathy in the first case. The long-read sequencing provided the strongest evidence in the process. Furthermore, PCR/Sanger sequencing provided exact information of the inversion breakpoints. The second patient performed the long-read sequencing directly and the identified large inversion mutation was confirmed by PCR/Sanger sequencing.

MATERIALS AND METHODS

RNA-Seq and cDNA capture sequencing

RNA-seq analysis was proven to be an effective method in diagnosing *DMD* variants [22]. Total RNA was extracted from muscle biopsy sample using the TRIzol reagent (Thermo Fisher), following the manufacturer's protocols. RNA integrity was evaluated by the Agilent 2100 Bioanalyzer (Agilent Technologies). Sample with RNA integrity number ≥7 were subjected to subsequent analysis. Libraries were constructed using VAHTS mRNA-seq v2 Library Prep Kit (Illumina) according to the manufacturer's instructions. The libraries were then sequenced on the Illumina sequencing platform (Illumina NovaSeq) with a sequencing strategy of 150 bp paired-end.

In the cDNA capture experiment, we first designed RNA probes of 120 bases in length (with a 60-base overlap) that cover all the cDNA regions of *DMD* gene, and then synthesized the probes by Agilent Technologies (Santa Clara, CA, USA). Hybridization of cDNA to the RNA probes was performed according to the protocols recommended for the SureSelect Target Enrichment system (Agilent). Purified DNA fragments were then subjected to sequencing on the Illumina sequencing platform (Illumina Hiseq X Ten). Paired-end reads were mapped to the human genome (GRCh38) using TopHat2 (2.1.1.) and htseqcount (0.11.1) was used to calculate the depth of exons of *DMD* gene.

Optical mapping

Optical mapping can facilitate direct examination of complex SVs, translocations, large insertions and inversions by creating more complete human diploid genome than the short-read sequencing methods [17]. For DNA extraction, 3 ml blood sample was used and high molecular weight genomic DNA was prepared by the SDS method followed by purification with QIAGEN® Genomic kit (Cat#13343, QIAGEN) according to the standard procedure provided by the manufacturer. The DNA degradation and contamination of the extracted DNA was monitored on 1% agarose gels. DNA purity was then detected using NanoDrop™ One UV-Vis spectrophotometer (Thermo Fisher Scientific, USA), of which OD260/280 ranging from 1.8 to 2.0 and OD 260/230 is between 2.0 and 2.2. At last, DNA concentration was further measured by Qubit® 3.0 Fluorometer (Invitrogen, USA). Optical sequencing was conducted based on the Bionano Saphyr platform as describe previously [23]. A standard automated pipeline for de novo assembly and SV calling was performed by Bionano Access v1.3 using human reference genome (GRCh38) to enable comprehensive SV analysis. The Bionano data has been deposited to a public database https://bigd.big.ac.cn/bioproject/browse/PRJCA002119.

Long-read DNA sequencing

Customized DNA probes of 100 bases were designed to cover DMD gene as well as its upstream and downstream 20 kb region (chrX:31115345-33377726). Probes corresponding to repetitive sequences in the human genome were excluded. Genomic DNA was extracted using the blood genome DNA extraction kit (Sangon Bioengineer Co., Shanghai, China) according to the manufacturer's protocol. The DNA quantity and quality was assessed with Qubit 3.0 (Thermo Fisher Scientific Inc., Carlsbad, CA, USA) and agarose gel electrophoresis, respectively. Targeted sequencing library was prepared as following: 3 µg genomic DNA per sample was sheared to 1–6 kb fragments by a g-Tube (#520079, Covaris) centrifugation (15,000 × *g*, 2 min, twice). The DNA was then purified after characterization of fragment size. End repair, A-tailing at the 3′ ends and adapter ligation was performed through pre-capture amplification. Targeted sequence capture was conducted by pooling indexed PCR products and hybridization with custom capture probes. Purified DNA fragments were amplified by PCR and quantified, then subjected to sequencing on the long-read sequencing platform Oxford Nanopore PromethION with R9.4.1 flow cell (for the first pedigree) and Pacific Biosciences with Sequel II platform (for the second pedigree) according to the manufacturer standard protocols. The sequences data have been deposited to a public database https://bigd.big.ac.cn/bioproject/browse/PRJCA002119.

Induced pluripotent stem cells and full-length cDNA sequencing

The ultra-pure (Purity > 95%) hiPSC-CMs were produced by Help Therapeutics and cryopreserved in Liquid Nitrogen. The cell preparation for epicardial injection was reported in previous paper (Guan X, Stem cell research and therapy, 2021). Briefly, Undifferentiated iPSCs grown to 90% confluence were induced to cardiomyocytes in a basal RPMI 1640 (Gibco) medium. On day 0 and day 1, 6 µM CHIR-99021 [Sigma-Aldrich] was supplemented. On day 3 and day 4, 5 µM of IWR-1 [Sigma-Aldrich] was supplemented subsequently. From day 8 and afterward, the medium was replenished with RPMI 1640 [Gibco] and B27 supplement [Gibco] every other day. Spontaneous contractions were noticed in between day 8 and 10 during differentiation. On day 16, purified iPSC-CMs were dissociated and 1 × 107 cells were cryopreserved per cryogenic tube for experiments.

Total RNA was extracted from iPSCs using the TRIzol reagent (Thermo Fisher) following the manufacturer's protocol. RNA integrity was evaluated using the Agilent 2100 Bioanalyzer (Agilent Technologies). Sample with RNA integrity number ≥7 were subjected to subsequent analysis. The cDNA-PCR Sequencing kit by Oxford Nanopore (SQK-PCS109) was used to prepare cDNA libraries. Briefly, 500 ng of total RNA, as measured by Qubit RNA HS assay (Thermo Fisher Scientific), was reverse transcribed using Superscript IV reverse transcriptase (Thermo Fisher Scientific). The cDNA was PCR amplified for 13 cycles with specific barcoded adapters from the Oxford Nanopore PCR Barcoding Kit (EXP-PBC001). The amplified cDNA was purified using AMPure XP beads (Beckman Coulter). Finally, the 1D sequencing adapter was ligated to the DNA then subjected to sequencing on Promethion platform using R9.4.1 flow cells according to the standard manufacturers conditions [24].

Sanger sequencing

Sanger sequencing were carried out to verify the long-read sequencing result in the two families. Particular front and back region of the breakpoints were used to design primers. The PCR was conducted under the following cycling conditions: 95 °C for 5 min, denaturation at 98 °C for 10 s, 10 cycles of amplification with annealing at 65 °C and another 25 cycles of amplification with annealing at 55 °C, with elongation at 72 °C for 40 s followed by a final elongation of 7 mins. The final PCR products were sequenced using Sanger sequencing and the Sanger sequence was aligned against GRCh19 reference genome to obtain the precise breakpoint information. The primers were provided in the supplementary information.

RESULTS

First pedigree

Conventional genetic tests failed to identify the pathogenic mutation in DMD*, while the patient exhibited the typical dystrophinopathy.* A male 5-year-old patient was referred to our hospital because of the gait abnormalities and frequent falls. The patient began to walk at age 2, but experienced difficulties in running, jumping, and climbing stairs. His physical examination showed calf and forearm muscle hypertrophy, weakness in pelvic-girdle muscles, a positive Gowers sign, and a waddling gait. The patient also had a definite family history (Fig. 1E): two of his maternal uncles and two first cousins suffered from the conditions similar to the patient: they lost their ambulation around 8–9 years old and died before 20 years old. Based on the information, we made a preliminary clinical diagnosis of his symptoms as dystrophinopathy and performed serum CK assay, which showed an elevated level of

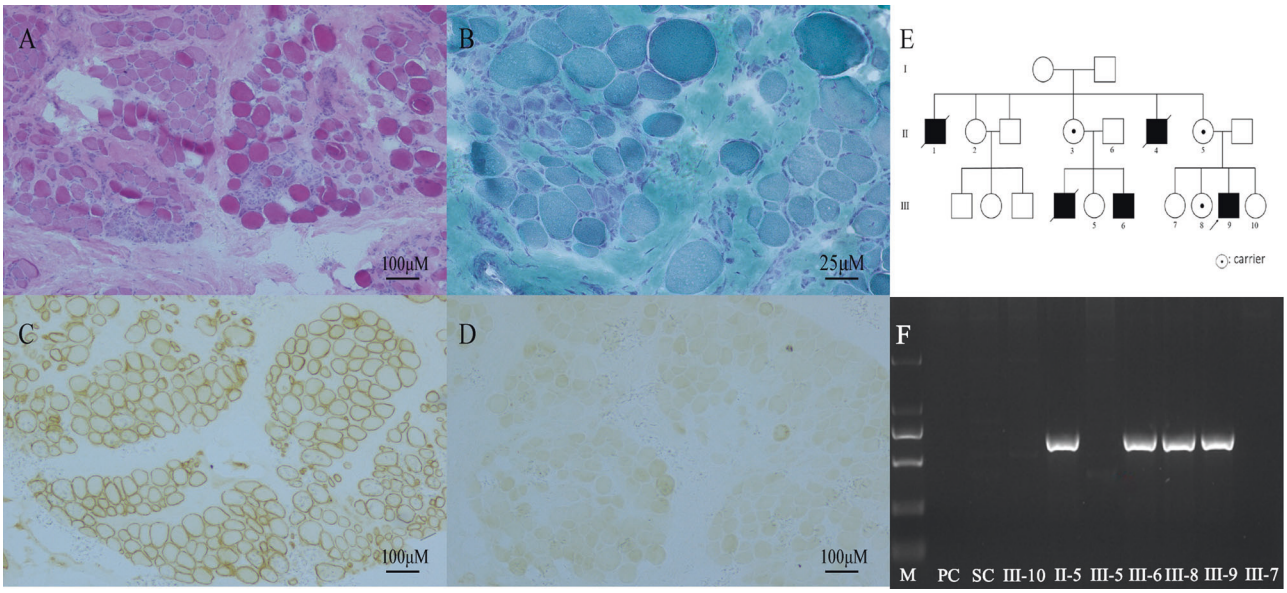

Fig. 1 The muscle pathology, family pedigree and diagnostic PCR results of the first pedigree. A Hematoxylin and Eosin staining revealed a typical dystrophic pattern (areas of fibrosis, fatty infiltrate, and degenerating and regenerating fibers). **B** Modified Gomori trichrome (MGT) staining showed a normal staining pattern. **C** Inmunohisto-chemistry (IHC) for a-sarcoglycan showed a normal staining pattern. **D** IHC for dystrophin showed the absence of expression. **E** The family pedigree of the patient.III-9: Proband, III-5: Unaffected first female cousin who was pregnant at the time of our findings. **F** The diagnostic PCR results of family members of the patient. Lane M: Marker; Lane PC: PCR control; Lance SC: sample control.

8100 U/L (normal: 55–300). Based on the clinical diagnosis of the pathophysiology, the proband has been treated with standard steroid treatment (prednisone 0.75 mg/kg/d) from July 2013 and been followed up regularly. Now he is 13-years old and can walk for a short distance independently.

To further validate our initial assessment of DMD, we performed genetic tests. However, conventional genetic tests failed to identify the pathogenic mutation in *DMD*, while the patient suffered from typical dystrophinopathy.

Firstly, we performed MLPA, which can detect deletions and duplications-the most common mutations in the *DMD* locus, but no deletion or duplication was detected. Next, we performed targeted WES using NGS to detect SNVs or splicing mutations in all of 79 exons of *DMD* gene, but no causative genetic mutation was identified (Supplementary Fig. 1A). Given the likely diagnosis of DMD, we performed a diagnostic muscle biopsy and immunohistochemistry tests, which revealed histological changes in muscle and absence of dystrophin expression in muscle fibers (Fig. 1A–D). Equipped with the definite molecular pathology results, we hypothesized there should be a rare type mutation at the *DMD* locus. Thus, we designed and carried out a secondary set of genetic tests for mutation(s) in *DMD*. At first, we performed karyotyping to determine whether there was a prominent chromosomal aberration that might drive the expression of the patient's phenotype. Unfortunately, the karyotyping also did not reveal any conspicuous rearrangement in chromosome X (Supplementary Fig. 1B).

RNA-Seq and cDNA capture sequencing analyses successfully identified loss of Dp427 transcripts, supporting the histological evidence for dystrophinopathy. We isolated RNA from the same muscle biopsy sample that was used for the histological study. We constructed libraries and performed RNA-seq to an average depth of ~150 million paired-end reads. The RNA-seq results suggested the absence of exon 3–55 transcripts and the presence of the transcripts of exons 1–2 and 56–79 (Fig. 2A). However, we considered that the results were rather inconclusive because of the relatively low overall sequencing depth (Fig. 2A).

Therefore, we further performed a cDNA capture experiment to obtain DNA sequences that showed high coverage and depth for individual exons. The captured cDNA showed a complete absence of the transcripts of exons 3–55 (Fig. 2B), while exon 1 transcript appeared to be intact based on the depth (20,000 copies) and coverage (99%), indicating the mutation might be present in exon 2 or between exon 2 and 3. The small amounts of transcripts covering exons 56–79 appear to be the products of the promoters that drive the expression of Dp116 and Dp71. The results from the cDNA capture experiment provided us the first clue for the causative genetic alterations.

Optical mapping identified the pathogenic inversion in DMD *locus: inv(X)(p21.1q21.1)*. Since the results of the cDNA capture indicated the possible presence of a complex SV such as balanced translocation or inversion at the *DMD* locus in the patient's genome, we decided to perform optical mapping to investigate the nature of the variant. A total of 294 Gb of sequence data at 95X depth coverage was obtained and an inversion event as the causative mutation in *DMD* locus in our patient genome was found (Fig. 3). The 55 Mb inversion was mapped between chrXp21.1 and chrXq21.1.

Long-read DNA sequencing precisely mapped the breakpoints of the inversion mutation. To gain insight in the discrepancy between targeted NGS and RNA-seq/cDNA capture results, we opted to use long-read DNA sequencing to effectively analyze the entire 2.2 Mb *DMD* gene. We generated a total of 55.23 Gb WGS ONT long reads with a read N50 of 19.12 kb. With 96.92% reads being mapped onto the reference genome, we were able to generate a depth coverage of 18.41 for the *DMD* gene. The breakpoints of the inversion were precisely mapped at the genomic locations of chrX:32,915,769 and chrX:87,989,329 (Fig. 4A). Detailed analyses of the breakpoints revealed the presence of SINE and LTR sequences at chrX:32,915,778-32,915,857 and chrX: 87,989,296-87,989,762, respectively (Supplementary Fig. 2). The 80 bp long SINE sequence locates at 10 bases downstream of the inversion breakpoint on Xp21.1 and the 467 bp long LTR element locates at the breakpoint

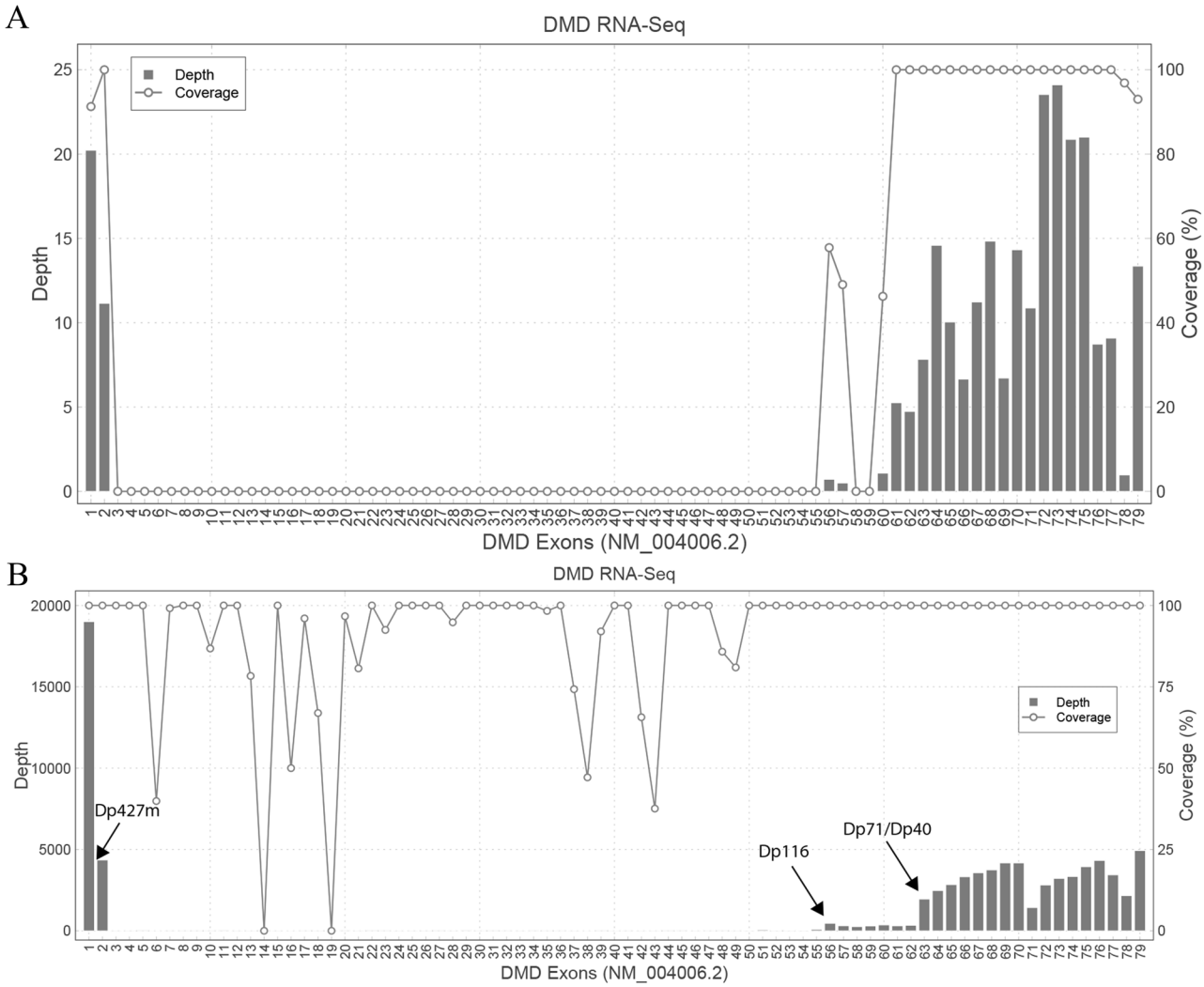

Fig. 2 The results of RNA sequencing and cDNA sequencing. A The result of RNA sequencing depth and coverage of RNA transcripts of exons of patient DMD gene, showing the missing exons 3–55. **B** The depth and coverage of DNA sequences of the captured cDNA exons, showing the absence of exons 3–55. Three different promoters (Dp427m, Dp116, Dp71/Dp40) are shown.

on Xq21.1. Similar results of special elements were found in the patient's breakpoint sequences. The inversion mutation appeared to prevent exons 3–55 from being transcribed.

Full-length cDNA sequencing identified abnormal transcripts generated by the inversion mutation. Considered that it has been 5 years since the patient received muscle biopsy and almost no muscle tissue remained, we further established iPSC from blood monocytes of the patient. The iPSC were differentiated into cardiomyocytes and full-length cDNA sequencing was performed to interrogate the transcript structure generated by the inversion mutation. We generated a total of 11,378,511 cDNA full-length reads with a read N50 of 1.58 kb. The four novel transcripts (Fig. 4B) detected in *DMD* included one transcript only composed of the first two exons of Dp427m, and other three transcripts that are the fusion transcripts between the first two exons of Dp427m transcript and the intergenic region in Xq21.1, which were consistent with the inversion mutation mentioned above.

New PCR-based genetic diagnosis was able to determine other family members' genotype. We designed a simple, PCR-based diagnostic test to detect the pathogenic inversion in *DMD* by using the precise sequence information at the inversion breakpoints that we obtained by nanopore sequencing (Supplementary Fig. 2). Since the inversion was enormous, full length of mutation could not be generated from one pair of primer. Therefore, two pairs of primers at each inversion breakpoint were designed. The details of the primers were summarized in Supplementary files. In addition, we identified a 16 bp novel insertion (AGTATTAAGTATTAAG) with microhomology in the patient DNA at the inversion breakpoints (Supplementary Fig. 3) by using Sanger sequencing. This PCR/Sanger sequencing results confirmed the same mutation in another patient in the family (the proband's male first cousin) and the carrier status for patient's mother, one of his sisters and one of his maternal aunts (Fig. 1F). Meanwhile, a female-first cousin of the patient was pregnant. Because of the family history, she was very concerned whether her child would be affected by the pathogenic mutation. Therefore, her DNA sample was tested. Fortunately, the PCR/Sanger sequencing result was negative for the inversion mutation, confirming her as a non-carrier (Fig. 1F). With this information, she made her own decision to carry her baby for the full-term delivery without fear.

Second pedigree

Basic information. The second patient was a 10-year-old boy. An elevated level of serum AST, ALT, CK and CK-MB was found by chance when the patient was at the age of 5 months. The patient then underwent MLPA but no deletion or duplication of *DMD*

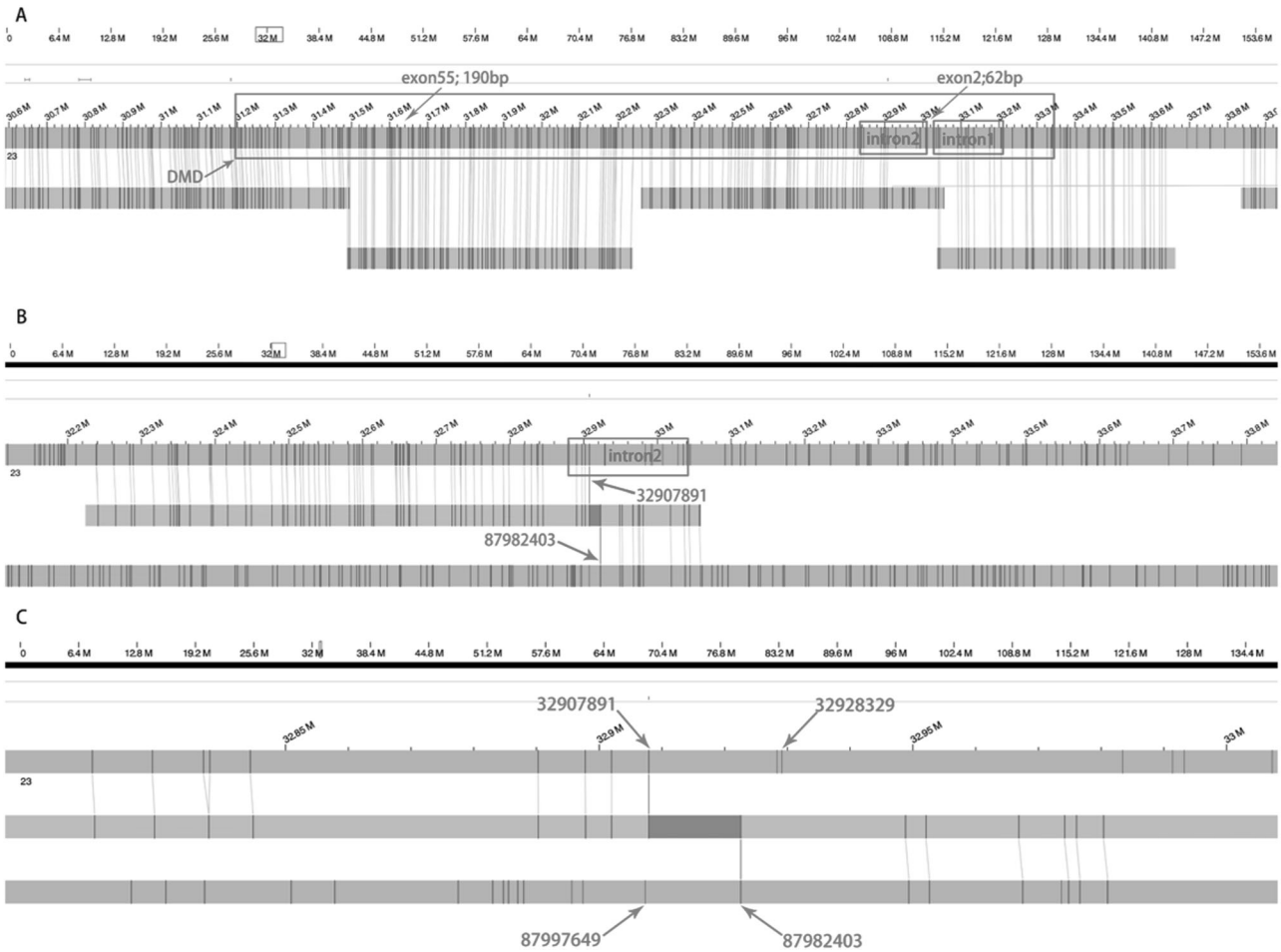

Fig. 3 The results of optical mapping of the pathogenic structural variation. The 55 Mb inversion was mapped between chrXp21.1 and chrXq21.31. **A** The mapped region of DMD gene that shows the inversion break points on exon 2. **B** The first break point is close to the mapping enzyme site at 32907891 and another is located close to the enzyme site at 87982403, indicating a potential pericentromeric inversion event. **C** The locations of mapping nucleases that span the putative break points.

exons was found. The patient began to walk at the age of 18 months, but was easy to fall and experienced difficulties in running, jumping, and climbing stairs. His physical examination showed a waddling gait and a positive Gowers' sign. And he underwent muscle biopsy at age of five, which showed the lack of dystrophin expression in muscle fibers. Based on the clinical characteristics and the muscle biopsy results, the patient was diagnosed with DMD. He has been treated with standard steroid treatment (prednisone 0.75 mg/kg/d) since age 6 and followed up until recently. At 2019, he received NGS but still failed to identify any copy number variations, SNVs or splicing mutations in *DMD* gene. Unfortunately, he showed a significant decrease of motor ability since July 2021, with more difficulties in walking, rising from squatting.

Long-read DNA sequencing directly found the breakpoint of large inversion mutation. Based on the previous findings, the second patient received long-read sequencing directly. The entire 2.2 Mb *DMD* gene was sequenced. A total of 296.31 Gb WGS ONT long reads with a read N50 of 3.26 kb was generated and 99.93% reads were mapped onto the reference genome. The median depth coverage of target region was 66.47. The breakpoints of the inversion were precisely mapped at the genomic locations of chrX: 32,581,576 and chrX: 127,797,236 (Fig. 4C). In the breakpoint junction, microhomology was observed. As for specific genetic elements, SINE, LINE and LTR were found near the breakpoints, as shown in Supplementary Fig. 4. Two SINEs, four LINEs and two LTRs were found in 1 kb near the two breakpoints. A LTR element was located 9 bp downstream of the distant-end breakpoint (chrX: 127,797,245-127797382). Similar results were found in the patient's breakpoint sequences.

PCR/Sanger sequencing confirmed the mutation found by long-read sequencing. PCR/Sanger sequencing was further conducted to confirm the result given by the long-read sequencing. Two primers were designed according to the long-read sequencing result for the positive and negative chains respectively. The details of the primers could be found in the Supplementary files. Based on Sanger sequencing, the inversion breakpoints were verified in the patient (Supplementary Fig. 5A, B). Besides, a 11 bp duplication of chrX: 127,797,226-127,797,236 (TCAGTCCTCTA) in the patient DNA was found (Supplementary Fig. 5A, B). The agarose electrophoresis proved that the proband and his mother had the mutation band, while the proband's father did not have the mutation band (Supplementary Fig. 6).

DISCUSSION

While conventional genetic tests have identified over 7000 mutations in *DMD* during the past three decades [6], new variants are still being found [19]. In our study, we report two new inversion variants that were not previously identified in *DMD*.

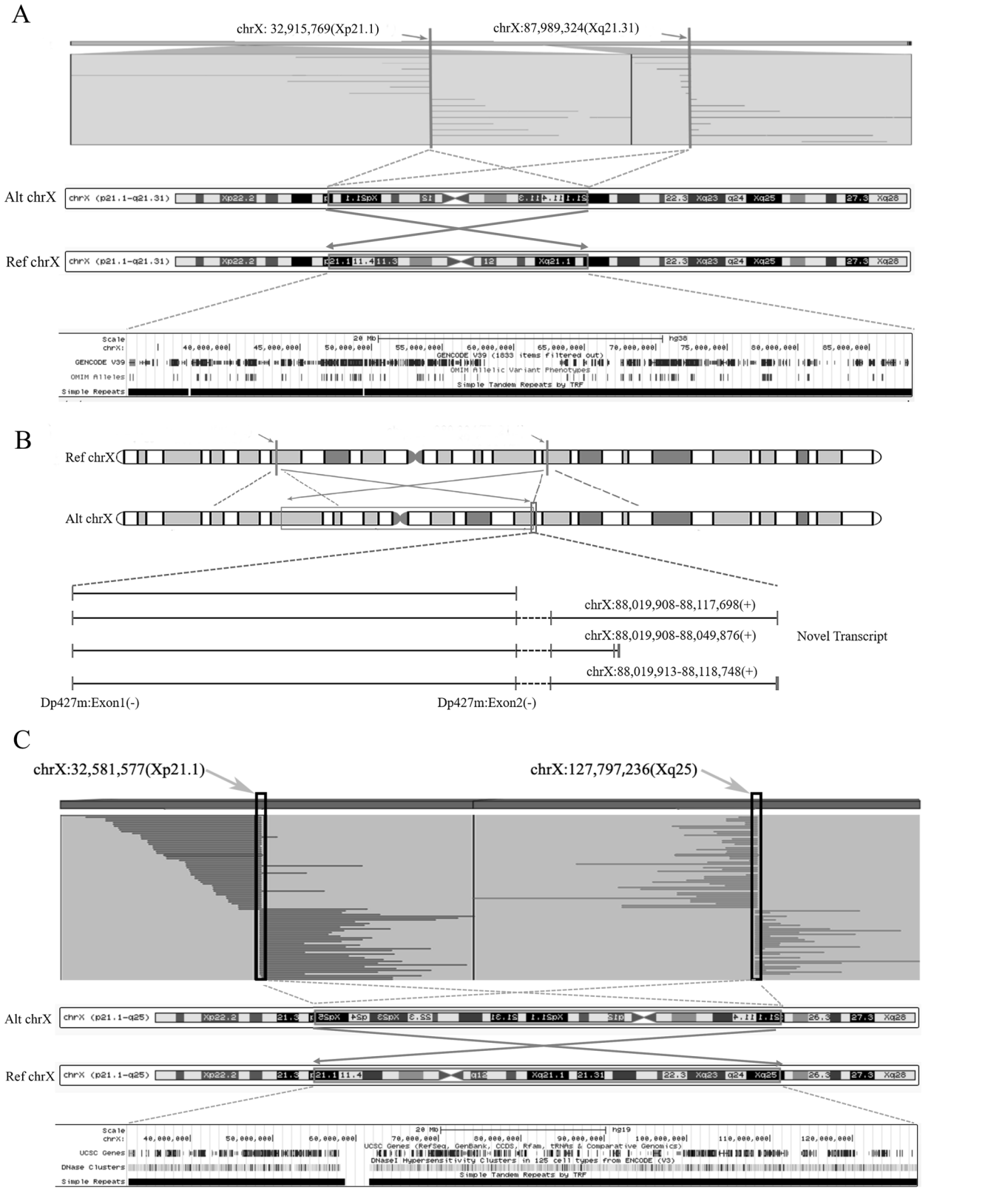

Fig. 4 The pathogenic inversion variant identified by long-read sequencing for two pedigrees. A The sequence reads that are precisely mapped to chrX:32915769(Xp21.1) and chrX:87989330 (Xq21.31), spanning 55 Mb across centromere on X chromosome. **B** A schematic picture of chromosome X with the inversion break points and novel transcript structure. For each transcript, boxes denote exons in black. **C** The sequence reads that are precisely mapped to chrX:32581577(Xp21.1) and chrX:127797236 (Xq25), spanning 95 Mb across centromere on X chromosome.

Because of the unique nature of the two variants, being large inversions more than 50 Mb encompassing the centromere of X chromosome, the majority conventional diagnosis methods failed to detect the mutations. In the first case, by building and applying an advanced pipeline, we successfully identified, mapped, and characterized the rare inversion *DMD* variant. Our pipeline consists of RNA-seq, cDNA capture, optical mapping, long-read sequencing and PCR/Sanger sequencing, while long-read sequencing was proved effective and played as a key role in this pipeline (Supplementary Fig. 7). This new pipeline proved powerful to detect new genetic variants that have not been identified by conventional genetic diagnosis methods such as MLPA, chromosome array, WES, short-read WGS, and karyotyping. Furthermore, in the second case, we then proved that long-read sequencing itself might be effective enough in diagnosis in complex SVs.

Although there are many reported mutations causing dystrophinopathy, it is still challenging to make a precise diagnosis of pathogenic variants for DMD patients. The challenges are due to the inherent limitations of the second-generation sequencing-based genetic tests. The PCR-based and short-read sequencing tests cannot identify genetic variants that are located at the genomic regions calcitrant to PCR amplification or sequencing. While short-read based WES and WGS are powerful to reveal point mutations, small insertions and deletions in intergenic and intronic regions [25, 26], they cannot accurately map the duplicated regions or structurally complex regions in the genome with highly repetitive sequences. Long-read sequencing, in contrast, is a novel and effective way to detect balanced translocations and inversions with high accuracy, while providing directionality and order information of SVs. It can also achieve satisfying mapping ratio in both exons and introns with repetitive sequences. Namely, long-read sequencing is more accurate in detecting inversions although in some cases short-read WGS might also be effective [27]. Therefore, it is essential to include long-read sequencing-based WGS and genome mapping technologies in diagnostic pipelines. In some cases, long-read sequencing might be directly conducted to detect rare mutations in conditions that traditional methods proved in vain.

The exact mechanism behind the origin of inversion mutations remains unclear. However, the repetitive sequences appear to be involved in the formation of the pathogenic inversion. There are SINE and LTR sequences at the breakpoints in the first case and LTR sequences at the breakpoints in the second case respectively. A variety of repetitive sequences are known to drive illegitimate recombination in the genome, resulting in the formation of SVs including inversion events [28, 29]. Furthermore, the 16 bp insertion sequence (AGTATTAAGTATTAAG) with repetitive ATTAAG in the first patient suggested that the double-strand break mediated non-homologous end joining could have been involved in the formation of the pathogenic inversion.

In the first pedigree the 55.0 Mb size inversion completely abolished the production of full-length Dp427 dystrophin polypeptides (Dp427b, Dp427m, Dp427p), causing the severe phenotype of the affected individuals. Based on the findings of DNA sequencing and mRNA resequencing, we verified the promotor of Dp427m and relevant elements initiate the transcription of exon 1, exon 2 and abnormally connected intergenic region between gene Kelch Like Family Member 4 (*KLHL4*) and CPX Chromosome Region Candidate 1 (*CPXCR1)*. The details remained unknown. As demonstrated by our comprehensive approach to identify the very challenging *DMD* variant, we might be able to determine any type of *DMD* variants with rigorous efforts.

Finally, based on the precise breakpoint information, we were able to design a PCR diagnostic method to confirm the presence of mutations in the first patient and to determine which members of the family are potential carriers of the variant. The assay determined one of his female-first cousins as a non-carrier and helped his cousin to make the decision to carry her baby for the full-term delivery of a healthy child.

Based on our results, in real-world practice, we suggest that patients who were clinically suspected with DMD/BMD (e.g., based on symptoms/serum CK/family history) should receive routine genetic tests (MLPA for CNVs or NGS), as the guideline recommends [30]. For those with no mutation found from conventional genetic test methods, current guidelines recommend muscle biopsy and RNA/cDNA sequencing [30]. However, long-read sequencing may be effective enough and less time consuming in detecting rare mutations like inversions, which can also provide the mutation information in DNA level and help physicians give genetic counseling suggestions to other family members. In some cases, we believe long-read sequencing might be performed directly to find atypical mutations. Optical mapping can help to map the locus in DNA yet could not provide the precise breakpoints information.

In conclusion, we have assembled an advanced SV analysis pipeline based on long-read sequencing to identify and characterize a DMD mutation that evaded detection by MLPA, WES and karyotyping. Combining cDNA capture, optical mapping, and long-read nanopore sequencing, we were able to identify a rare and large inversion variant of *DMD*. Besides, direct use of long-read sequencing also successfully detected another inversion variant in the second case. Considering that it's less time consuming and can provide precise breakpoint information, long-read sequencing may possess significant value in DMD diagnosis in future.

DATA AVAILABILITY

The data are available from the corresponding author on reasonable request.

REFERENCES

1. Burghes AH, Logan C, Hu X, Belfall B, Worton RG, Ray PN. A cDNA clone from the Duchenne/Becker muscular dystrophy gene. Nature. 1987;328:434–7.
2. Kunkel LM, Monaco AP, Middlesworth W, Ochs HD, Latt SA. Specific cloning of DNA fragments absent from the DNA of a male patient with an X chromosome deletion. Proc Natl Acad Sci USA. 1985;82:4778–82.
3. Emery AE. The muscular dystrophies. Lancet 2002;359:687–95.
4. Aartsma-Rus A, Van Deutekom JC, Fokkema IF, Van Ommen GJB, Den Dunnen JT. Entries in the Leiden Duchenne muscular dystrophy mutation database: An overview of mutation types and paradoxical cases that confirm the reading-frame rule. Muscle Nerve: Off J Am Assoc Electrodiagn Med. 2006;34:135–44.
5. Rybakova IN, Patel JR, Ervasti JM. The dystrophin complex forms a mechanically strong link between the sarcolemma and costameric actin. J Cell Biol. 2000;150:1209–14.
6. Koenig M, Beggs A, Moyer M, Scherpf S, Heindrich K, Bettecken T, et al. The molecular basis for Duchenne versus Becker muscular dystrophy: correlation of severity with type of deletion. Am J Hum Genet. 1989;45:498.
7. Tuffery-Giraud S, Béroud C, Leturcq F, Yaou RB, Hamroun D, Michel-Calemard L, et al. Genotype–phenotype analysis in 2,405 patients with a dystrophinopathy using the UMD–DMD database: a model of nationwide knowledgebase. Hum Mutat. 2009;30:934–45.
8. Bladen CL, Salgado D, Monges S, Foncuberta ME, Kekou K, Kosma K, et al. The TREAT-NMD DMD Global Database: analysis of more than 7,000 Duchenne muscular dystrophy mutations. Hum Mutat. 2015;36:395–402.
9. Tong YR, Geng C, Guan YZ, Zhao YH, Ren HT, Yao FX, et al. A Comprehensive Analysis of 2013 Dystrophinopathies in China: A Report From National Rare Disease Center. Front Neurol. 2020;11:572006.
10. Trippe H, Wieczorek S, Kötting J, Kress W, Schara U. Xp21/A translocation: a rarely considered genetic cause for manifesting carriers of duchenne muscular dystrophy. Neuropediatrics 2014;45:333–5.
11. Ligon AH, Kashork CD, Richards CS, Shaffer LG. Identification of female carriers for Duchenne and Becker muscular dystrophies using a FISH-based approach. Eur J Hum Genet. 2000;8:293–8.
12. Ishmukhametova A, Van Kien PK, Méchin D, Thorel D, Vincent M-C, Rivier F, et al. Comprehensive oligonucleotide array-comparative genomic hybridization analysis: new insights into the molecular pathology of the DMD gene. Eur J Hum Genet. 2012;20:1096–100.

13. van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. The Third Revolution in Sequencing Technology. Trends Genet. 2018;34:666–81.
14. Zhang Q, Xu X, Ding L, Li H, Xu C, Gong Y, et al. Clinical application of single-molecule optical mapping to a multigeneration FSHD1 pedigree. Mol Genet Genom Med. 2019;7:e565.
15. Roberts RJ, Carneiro MO, Schatz MC. The advantages of SMRT sequencing. Genome Biol. 2013;14:405.
16. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat Methods. 2015;12:733–5.
17. Barseghyan H, Tang W, Wang RT, Almalvez M, Segura E, Bramble MS, et al. Next-generation mapping: a novel approach for detection of pathogenic structural variants with a potential utility in clinical diagnosis. Genome Med. 2017;9:90.
18. Mantere T, Kersten S, Hoischen A. Long-read sequencing emerging in medical genetics. Front Genet. 2019;10:426.
19. Chaisson MJ, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, et al. Resolving the complexity of the human genome using single-molecule sequencing. Nature 2015;517:608–11.
20. Sanchis-Juan A, Stephens J, French CE, Gleadall N, Mégy K, Penkett C, et al. Complex structural variants in Mendelian disorders: identification and breakpoint resolution using short-and long-read genome sequencing. Genome Med. 2018;10:95.
21. Collins RL, Brand H, Redin CE, Hanscom C, Antolik C, Stone MR, et al. Defining the diverse spectrum of inversions, complex structural variation, and chromothripsis in the morbid human genome. Genome Biol. 2017;18:36.
22. Gonorazky H, Liang M, Cummings B, Lek M, Micallef J, Hawkins C, et al. RNA seq analysis for the diagnosis of muscular dystrophy. Ann Clin Transl Neurol. 2016;3:55–60.
23. Dai Y, Li P, Wang Z, Liang F, Yang F, Fang L, et al. Single-molecule optical mapping enables quantitative measurement of D4Z4 repeats in facioscapulo-humeral muscular dystrophy (FSHD). J Med Genet. 2020;57:109–20.
24. Deng J, Yu J, Li P, Luan X, Cao L, Zhao J, et al. Expansion of GGC Repeat in GIPC1 Is Associated with Oculopharyngodistal Myopathy. Am J Hum Genet. 2020;106:793–804.
25. Okubo M, Minami N, Goto K, Goto Y, Noguchi S, Mitsuhashi S, et al. Genetic diagnosis of Duchenne/Becker muscular dystrophy using next-generation sequencing: validation analysis of DMD mutations. J Hum Genet. 2016;61:483–9.
26. 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.
27. Falzarano MS, Grilli A, Zia S, Fang M, Rossi R, Gualandi F, et al. RNA-seq in DMD urinary stem cells recognized muscle-related transcription signatures and addressed the identification of atypical mutations by whole-genome sequencing. HGG Adv. 2022;3:100054.
28. Weckselblatt B, Rudd MK. Human structural variation: mechanisms of chromosome rearrangements. Trends Genet. 2015;31:587–99.
29. Carvalho CM, Ramocki MB, Pehlivan D, Franco LM, Gonzaga-Jauregui C, Fang P, et al. Inverted genomic segments and complex triplication rearrangements are mediated by inverted repeats in the human genome. Nat Genet. 2011;43:1074.
30. Fratter C, Dalgleish R, Allen SK, Santos R, Abbs S, Tuffery-Giraud S, et al. EMQN best practice guidelines for genetic testing in dystrophinopathies. Eur J Hum Genet. 2020;28:1141–59.

ACKNOWLEDGEMENTS

The authors express sincere gratitude to the patient and the family members who participated in the study. Special acknowledgements go to Dr Xiaoming (BGI Clinical Laboratory, Wuhan, China) for assistance for muscle tissue mRNA sequencing and to Ms Minming Zheng (Grandomics Biosciences, Beijing, China) for preparing the figures.

AUTHOR CONTRIBUTIONS

YD and JP designed the study. CG, CZ, YD, and JP collected clinical information of the pedigree. BZ and JH performed karyotype and PCR tests of family members. YZ and LC performed the pathology analysis of the muscle biopsy. PL, FL, and HJ conducted Bionano and ONT data analysis, YW and YRW performed Bionano and ONT sequencing and data collection, FY performed MLPA experiment. YD, JP LYC, DL, SL, DW, and LC supervised the project. CG, CZ, YT, MSP, YD, and JP drafted the paper with input from all authors.

FUNDING

This study was funded by CAMS Innovation Fund for Medical Sciences (CIFMS) (No.2016-I2M-1-002) and Chinese National Basic Research Program (973) (No.2017YFC1001902).

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS APPROVAL

This study was approved by the ethics committee of the Peking Union Medical College Hospital (IRB #JS-1233). All participants and the parents of those <18years of age at the time of genetic testing gave written informed consent.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41431-022-01190-y.

Correspondence and requests for materials should be addressed to Jing Peng or Yi Dai.