

RESEARCH ARTICLE

Exonic rearrangements in *DMD* in Chinese Han individuals affected with Duchenne and Becker muscular dystrophies

Chao Ling¹  | Yi Dai² | Li Fang³ | Fengxia Yao¹ | Zhe Liu¹ | Zhengqing Qiu⁴ | Liying Cui² | Fan Xia⁵ | Chen Zhao⁵ | Shuyang Zhang⁶ | Kai Wang^{3,7} | Xue Zhang^{1,8}

¹The Laboratory of Clinical Genetics, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China

²Department of Neurology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China

³Department of Pathology & Laboratory Medicine, Raymond G. Perelman Center for Cellular and Molecular Therapeutics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania

⁴Department of Pediatrics, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China

⁵Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas

⁶Department of Cardiology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China

⁷Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania

⁸Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, McKusick-Zhang Center for Genetic Medicine, School of Basic Medicine Peking, Union Medical College, Beijing, China

Correspondence

Kai Wang, Raymond G. Perelman Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, PA 19104.
Email: wangk@email.chop.edu

Xue Zhang, Laboratory of Clinical Genetics, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Peking Union Medical College, 100730 Beijing, China.
Email: xuezhang@pumc.edu.cn

Funding information

Chinese Academy of Medical Sciences (CAMS), Initiative for Basic Scientific Research, Grant/Award Numbers: 2015PT320017, 2016RC310006

Abstract

Exonic deletions and duplications within *DMD* are the main pathogenic variants in Duchenne and Becker muscular dystrophies (DMD/BMD). However, few studies have profiled the flanking sequences of breakpoints and the potential mechanism underlying the breakpoints in different fragile regions of *DMD*. In this study, 896 Chinese male probands afflicted with DMD/BMD were selected from unrelated families and analyzed using multiplex ligation-dependent probe amplification of the *DMD* gene, in which we identified exon deletions in 784 subjects and duplications in 112 subjects. Deletions occurred most frequently in the genomic region encompassing exons 45–55, accounting for 73% of all deletion patterns. Furthermore, to unravel the potential mechanism that induced breaks, *DMD* gene capture and sequencing were performed to identify the breakpoints in 37 subjects with deletions encompassing exons 45–55 of *DMD*; we found that *DMD* instability did not arise from a single cause; instead, long-sequence motifs, nonconsensus microhomologies, low-copy repeats, and microindels were embedded around the breakpoints, which may predispose *DMD* to instability. In summary, this study highlights the heterogeneous characteristics of the flanking sequences around the breakpoints and helps us to understand the mechanism underlying *DMD* gene instability.

KEYWORDS

breakpoints, *DMD* gene instability, flanking sequences, long-sequence motif, recombination

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2019 The Authors. *Human Mutation* Published by Wiley Periodicals, Inc.

1 | INTRODUCTION

Duchenne and Becker muscular dystrophies (DMD/BMD) are X-linked recessive disorders characterized by progressive muscle degeneration and weakness and are caused by variants in the *DMD* gene (Tuffery-Giraud et al., 2009). Affected subjects are primarily males and show early onset of symptoms. Previous studies have described the *DMD* gene variation spectrum, which includes deletions/duplications, small rearrangements, and point mutations. More than 70% of diagnosed patients show large deletions in the *DMD* gene, while more than 10% show large duplications (Tuffery-Giraud et al., 2009). Several studies have reported that the major deletion hotspots are around exons 45–52 and 8–13 and that the deletion patterns differ among populations (Onengut et al., 2000). Although different patients may have the same deletion, their breakpoints can differ (Nobile et al., 2002). Multiplex ligation-dependent probe amplification (MLPA) is commonly used for the detection of large-scale *DMD* exon deletions and duplications, but the mechanisms underlying multiexonic deletions or duplications have not been investigated in-depth due to the technical limitations of whole *DMD* gene sequencing: the breakpoints occur mainly in the introns, which are too large to be sequenced using conventional sequencing techniques. Previous genomic instability studies have revealed common fragile sites (CFSs) corresponding to the regions showing genomic rearrangements and displaying delayed replication (Casper, Nghiem, Arlt, & Glover, 2002). These sites might be the result of the nucleotide sequences and/or chromosomal structures. Specifically, microhomologies are predominantly involved in rearrangement processes (Mitsui et al., 2010). However, detailed information about deletion-prone regions could not be elucidated, and some of the CFSs could not be explained under the current mechanism (Ishmukhametova et al., 2012). Thus, to unravel the potential mechanism(s) that induced breaks in the *DMD* gene, we performed an in-depth analysis of the deletions and duplications in 896 DMD/BMD male probands. Additionally, 37 patients were scanned using a specific whole *DMD* gene panel and sequenced using next-generation sequencing (NGS). Long-sequence motifs in areas surrounding the breakpoints were investigated to assist in our understanding of the potential mechanism of *DMD* instability.

2 | MATERIALS AND METHODS

2.1 | Subjects

A total of 896 Chinese Han male probands from unrelated families who were admitted to Peking Union Medical College Hospital from 2007 to 2017 were selected for the present study. These probands had exon deletions or duplications in the *DMD* gene as determined by MLPA (MRC-Holland, The Netherlands) and were definitely diagnosed as DMD/BMD (Bushby et al., 2010). To discover the potential mechanism underlying the *DMD* break, 37 patients were enrolled in this study for whole-gene capture-based sequencing, including 35

patients with deletions encompassing exons 45–55 and two relatively low-frequency deletions of exons 7–41 and exons 63–64, which occurred up- and downstream of exons 45–55 region of *DMD*, respectively. These patients were used to determine whether the up- and downstream regions had the same characteristics as the common deletion region. This study was approved by the duly constituted ethics committee of Peking Union Medical College Hospital, and all of the subjects were anonymized for sequencing data analysis.

2.2 | Multiplex ligation-dependent probe amplification

The deletion and duplication patterns were identified using MLPA using a kit (MRC-Holland, The Netherlands). The kit includes two probe mixes, P034-B2 and P035-B1, which together contained one probe for each of the exons on the Xp21.2 chromosome of the *DMD* gene (79 exons). In addition, one probe was present in P035-B1 for the alternative exon 1 found in the transcript variant Dp427c of *DMD*. Thus, performing two MLPA reactions was sufficient to investigate the copy number of all exons. The MLPA experiments were performed according to the manufacturer's instructions, and the ABI 3130XL and 3700XL genetic analyzers were used for fragment separation and analysis. In addition, the single exon deletion in the *DMD* gene was further validated using multiplex-polymerase chain reaction and agarose electrophoresis methods.

2.3 | Whole *DMD* gene capture and sequencing

For 37 patients, genomic DNA was sequenced using the whole *DMD* gene capture technique (NGS+MyGenoCap; MyGenostics). Biotinylated single-strand DNA capture probes were used to capture the whole *DMD* gene. DNA samples were prepared as Illumina sequencing libraries, which were enriched for the *DMD* gene with the MyGenostics target region enrichment protocol. The captured libraries were sequenced using an Illumina NextSeq 500 sequencer. The capture efficiency was over 99%, and the average sequencing depth was 200 times in the normal *DMD* gene. In this study, the deleted exons could not be captured and amplified, so the DMD/BMD subjects had a relatively low capture percentage of ~83% when we mapped the reads to the complete *DMD* gene. The average sequencing depth of the target was 148 times. When focusing only on exonic regions, over 92% of the exonic regions in *DMD* was captured in sequence data, and the average coverage depth was 190 times. FASTQ files were extracted, and Illumina sequencing adaptors, sort reads, and low-quality reads were filtered by Cutadapt software (Martin, 2011). Quality control analysis was performed using the FastQC pipeline (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Clean reads were mapped to the human reference genome (GRCh37/hg19) using Burrows-Wheeler Aligner software (Li & Durbin, 2010). The soft-clipped reads were extracted with CREST (Wang et al., 2011) to analyze multiexonic deletion variations. Integrative genomics viewer was used for breakpoint identification (Robinson et al., 2011; Thorvaldsdottir, Robinson, & Mesirov, 2013).

2.4 | Flank reads analysis

After accurate breakpoints were identified, the reads flanking each breakpoint were selected. There were two breakpoints in each of the 37 samples, which we named the 5-prime breakpoint and the 3-prime breakpoint. To discover the novel ungapped motifs, Multiple Em for Motif Elicitation software (MEME, version 4.12.0) was used. Based on a previous study (Mitsui et al., 2010) and the default maximum motif width (50 base pair [bp]), we selected 200 bp flanking reads up- and downstream of the 5-prime breakpoint, and the same method was used for the 3-prime breakpoint in all 37 subjects (Figure S1). Therefore, a total of 74 reads were uploaded to the MEME website, and “zero or one occurrence per sequence” was selected as the occurrences of motifs among the sequences. MEME uses statistical modeling techniques (using a heuristic function) to automatically choose the best width, with a maximum of 50 bp. The number of occurrences and the description of each motif were outputted for this study.

2.5 | Sanger sequencing

In this study, a total of 74 breakpoints were validated using Sanger sequencing. The forward primer was designed with the upstream sequence of the 5-prime breakpoint, and the reverse primer was designed with the downstream sequence of the 3-prime breakpoint using Primer3web (version 4.0.0) software (Table S4). TaKaRa LA Taq was used for PCR following the manufacturer’s protocol. All of the identified variations were submitted to the <http://www.LOVD.nl/DMD> database (Table S3).

3 | RESULTS

3.1 | Spectrum of large-scale variations in the *DMD* gene

Over an approximately 10-year period, we recruited a total of 896 MLPA confirmed DMD/BMD probands from unrelated families for the present study. At the age of definite diagnosis, we found no significant difference between the deletion (7.3 ± 0.2 ; $n = 626$) and duplication (8.3 ± 0.75 ; $n = 93$) patients; however, there was an obvious age difference ($p < .0001$) between the patients with DMD (6.97 ± 0.18 ; $n = 593$ available) and the patients with BMD (9.71 ± 0.85 ; $n = 127$ available; Table S3). We further analyzed variant patterns of both the exon deletions (784 patients) and duplications (112 patients). A profile of the deletion patterns observed in the 784 patients with DMD/BMD is shown in Figure 1, and we found that deletions primarily occurred in the central region of the *DMD* gene. Detailed statistics revealed that the deletion of exons 48–50 occurred most frequently (6.76%, 53/784) in all of the DMD/BMD deletion cases (Figure S2; Table S1). Analysis of the breakpoint distribution around deletion-prone regions illustrated that the 5-prime breakpoint mostly occurred in intron 44 ($n = 173/784$), while the 3-prime breakpoint mostly occurred in intron 50 ($n = 154/784$),

and the whole breakpoint (including both of the 5-prime and the 3-prime) distribution analysis illustrated that intron 44 ($n = 212/784$) was the most frequent break region, followed by intron 50 ($n = 198$; Figure S3a–c; Table S2). Additionally, in the subjects with exon duplication, we observed that exon 2 duplication ($n = 10$) occurred the most frequently, followed by duplication in exons 53–55 (Figure S4; Table S1).

Additionally, we separated the BMD ($n = 141$) cases from the DMD cases ($n = 643$), and the deletion pattern and breakpoint distribution analysis uncovered that the deletion of exons 45–47 was the most common (2.20%, 31/141) in patients with BMD, and the breakpoint was commonly in intron 44 (52.48%, 74/141; 5-prime, $n = 62$; 3-prime, $n = 12$). The most frequent deletion pattern in patients with DMD was exons 48–50 (8.24%, 53/643), and the breakpoint distribution mostly occurred in intron 50 (30.79%, 198/643; 5-prime, $n = 44$; 3-prime, $n = 154$), followed by intron 45 (22.55%, 145/643; 5-prime, $n = 40$; 3-prime, $n = 145$) and intron 44 (21.46%, 138/643; 5-prime, $n = 111$; 3-prime, $n = 27$; Table S2).

3.2 | Breakpoint detection of the *DMD* gene

The exact breakpoints in *DMD* genomic DNA were analyzed in 37 selected subjects using a custom in-solution capture panel, which was designed for the whole *DMD* gene, and the samples were subjected to NGS. Breakpoints were identified and subsequently validated using Sanger sequencing with the designed primers (Table 1; Table S4). The sequencing results indicated that the exact breakpoints were all located in *DMD* introns, and the same breakpoint (c.6615-1512_7660+21124del) was observed in two unrelated patients with the same deletion pattern (exons 46–52 deletions). However, the other breakpoints were all unique, although some of them resulted in the same deletion pattern. We further performed reading-frame analysis with *DMD* reading-frame checker 1.9 software (http://www.humgen.nl/scripts/DMD_frame.php) on all cases. Here, 161 of the 896 cases were predicted to have in-frame mutations, and 735 of them were predicted to have out-of-frame mutations. The reading frame analysis was consistent with the BMD in-frame and DMD out-of-frame rules (Table 1; Table S3).

3.3 | *DMD* gene instability analysis

To identify the characteristics of the break-prone regions in the *DMD* gene and to investigate the potential mechanisms underlying *DMD* breaks, we isolated a 200 bp read up- and downstream of the breakpoints. Manual microhomology analysis was first performed between the paired flanking sequences. Briefly, 20/37 patients had junctions with microhomologies; 4/37 patients had extended microhomologies; 5/37 patients had inserted sequences; and the breakpoints in 7/37 patients occurred mostly in the low copy repeat region. In all, the microhomologies and the extended microhomologies accounted for ~65% (24/37) of the intragenic recombination in the *DMD* gene (Table 2). However, there was still no detailed information on the remaining nonhomologous end-joining sequences.

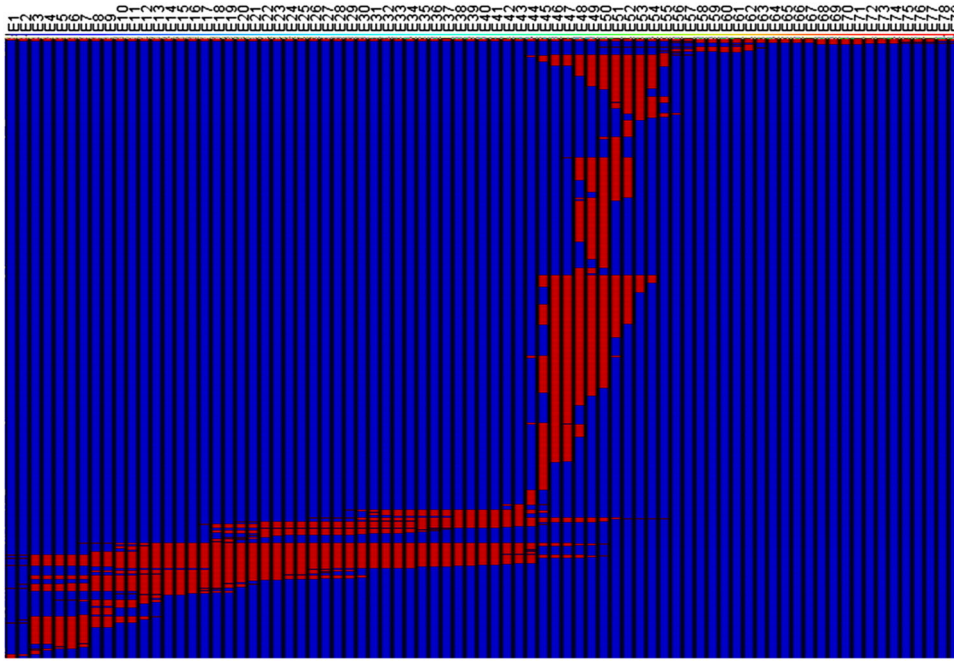


FIGURE 1 The exon deletion patterns for 784 patients with DMD/BMD. A multiple array viewer was used to make the deletion pattern cluster. The 79 exons of *DMD* are listed in order, and the red bars represent the deleted exons, and the blue bars indicate the normal exons. BMD, Becker muscular dystrophy; DMD, Duchenne muscular dystrophy

Thus, MEME was used for long-sequence motif analysis (Bailey et al., 2009). To capture more sufficient information, we used 200-bp flanking reads and performed the motif analysis as described in Section 2. Here, we identified two motifs of 49 bp ($p < 2.55e10^{-19}$) and 41 bp ($p < 6.10e10^{-19}$). The 49 bp motif (motif A), which was predicted to be a spliceosomal complex, was observed in 7/37 patients with DMD/BMD (Figure 2a), while the 41 bp motif (motif B), predicted to be involved in nucleotide and nucleic acid metabolic processes, was observed in 10/37 patients with DMD/BMD (Figure 2b). Additionally, 5/37 of the individuals, who had no microhomologies, had the predicted long motifs around the breakpoints. Furthermore, six patients shared both long motifs (Table 2). We inspected long motif A and motif B and found that they appeared to map to AluS and AluY elements. We predicted that the sequences repeated in the *DMD* gene might act as Alu elements and interfere with DNA replication, recombination, and repair, which would make these sequences the main cause of *DMD* instability.

4 | DISCUSSION

In the present study, we selected 896 unrelated DMD/BMD male probands with deletion/duplication detected by MLPA to characterize the multiexonic deletion or duplication variations in *DMD*. In 784 patients affected with deletion, we found that the pattern involving exons 48–50 occurred most frequently, followed by that involving exons 45–50 and exon 45 (Figure S2; Table S1). However, when we separated the 141 patients with BMD from the patients with DMD, we found that the deletion of exons 45–47 most frequently occurred in

patients with BMD. A recent study of 317 patients with DMD/BMD from southern India showed that the most frequent deletion was in exons 45–47 (Vengalil et al., 2017). Deletion of exons 45–47 was also the most frequently occurring pattern in 141 patients with DMD/BMD from Puerto Rico (Ramos et al., 2016). In 180 Polish patients with DMD/BMD, the deletion mainly involved exons 45–54 and exons 3–21 (Zimowski et al., 2014). In 1,497 Japanese patients with DMD/BMD, exon deletions were most frequently observed in the central hot spot region between exons 45 and 52 (Okubo et al., 2017), which was consistent with the results of this study (Figure 1). Thus, the *DMD* gene deletion spectrum observed here was similar to that reported in patients with DMD/BMD in Asia, but slightly different from that in patients in Europe and America, which might be due to the limited sample size. Additionally, the different deletion patterns in patients with DMD and BMD could contribute to the whole-variant pattern when there is a relatively small sample size.

We further investigated the potential mechanisms underlying exonic deletion variations in the *DMD* gene through gene capture and NGS at single-base resolution in 37 selected patients. We observed that the breakpoints occurred in all introns of the *DMD* gene; however, they occurred more frequently from intron 44 to intron 54, here accounting for 73% of *DMD* intragenic deletions in our DMD/BMD subjects, and intron 44 was found to be the most frequent break region in this cohort (Figure S3a–c). The selected flanking reads analysis in intron 44 found that microhomology or extended microhomology sequences existed around the junction reads, and one subject (D37) without microhomology sequences had long motifs (Tables 1 and 2). The above results illustrate that microhomology (Marey et al., 2016) and long motifs might contribute to the

TABLE 1 Patient information and the breakpoints within the *DMD* gene

ID	Gender	Age at diagnosis	Locus Reference Genomic (LRG_199t1 references)	Deletion region	Reading frame	Phenotype
D1	Male	8	c.7201-4001_7660+13394del	Introns 49–52	Out-of-frame	DMD
D2	Male	9	c.9287-2941_9361+940delinsTAG	Introns 63–64	In-frame	BMD
D3	Male	9	c.7099-1901_8027+8827del	Introns 48–54	Out-of-frame	DMD
D4	Male	15	c.6615-15465_8547+5882del	Introns 45–57	Out-of-frame	DMD
D5	Male	4	c.8028-2087_9163+153delinsATAACTTG	Introns 54–61	Out-of-frame	DMD
D6	Male	6	c.6614+13881_7098+13715del	Introns 45–48	Out-of-frame	DMD
D7	Male	3	c.650-53656_5922+6505delinsAT	Introns 7–41	Out-of-frame	DMD
D8	Male	3	c.7099-10208_7661–22210del	Introns 48–52	Out-of-frame	DMD
D9	Male	12	c.6615-11775_7309+4771del	Introns 45–50	Out-of-frame	DMD
D10	Male	8	c.7099-12971_8027+3726del	Introns 48–54	Out-of-frame	DMD
D11	Male	10	c.7098+12114_8027+7417del	Introns 48–54	Out-of-frame	DMD
D12	Male	3	c.7309+5104_7542+17062del	Introns 50–51	Out-of-frame	DMD
D13	Male	3	c.6614+5702_8217+43627del	Introns 45–55	Out-of-frame	DMD
D14	Male	3	c.6912+18064_8027+2675del	Introns 47–54	Out-of-frame	DMD
D15	Male	3	c.7098+7776_7660+1050del	Introns 48–52	Out-of-frame	DMD
D16	Male	3	c.7309+15311_7872+6382del	Introns 50–53	Out-of-frame	DMD
D17	Male	15	c.6438+102544_6913-15866del	Introns 44–47	In-frame	BMD
D18	Male	7	c.7309+14745_8028-7983del	Introns 50–54	Out-of-frame	DMD
D19	Male	11	c.7099-10048_8028-4789delinsCAC	Introns 48–54	Out-of-frame	DMD
D20	Male	2	c.7309+4829_8027+1051del	Introns 50–54	Out-of-frame	DMD
D21	Male	9	c.6912+3306_8027+6905del	Introns 47–54	Out-of-frame	DMD
D22	Male	11	c.7201-7674_7310-20273del	Introns 49–50	Out-of-frame	DMD
D23	Male	16	c.6439-28736_6913-10868del	Introns 44–47	In-frame	BMD
D24	Male	5	c.7310-11092_8218-59513del	Introns 50–55	Out-of-frame	DMD
D25	Male	3	c.6614+13221_7310-2598delinsT	Introns 45–50	Out-of-frame	DMD
D26	Male	7	c.6913-6367_7661-17902del	Introns 47–52	Out-of-frame	DMD
D27	Male	4	c.7099-1406_7660+5892del	Introns 48–52	Out-of-frame	DMD
D28	Male	4	c.6615-1512_7660+21124del	Introns 45–52	Out-of-frame	DMD
D29	Male	8	c.6615-1512_7660+21124del	Introns 45–52	Out-of-frame	DMD
D30	Male	3	c.6439-49382_7872+308del	Introns 44–53	In-frame	BMD
D31	Male	3	c.6913-9844_7310-16920del	Introns 47–50	Out-of-frame	DMD
D32	Male	8	c.6615-419_7200+5420del	Introns 45–49	Out-of-frame	DMD
D33	Male	7	c.7543-14415_8027+1540del	Introns 51–54	Out-of-frame	DMD
D34	Male	2	c.6439-40173_7661-320del	Introns 44–52	Out-of-frame	DMD
D35	Male	3	c.6439-97179_7200+8061del	Introns 44–49	In-frame	BMD
D36	Male	2	c.6439-48962_7310-2533del	Introns 44–50	Out-of-frame	DMD
D37	Male	4	c.6439-16684_8217+15910del	Introns 44–55	In-frame	BMD

Note: The basic information of the 37 selected patients is listed in the table. The breakpoints and the intron locations were presented, and the GRCh37/hg19, LRG_199t1 references were used in the sequencing data analysis. In addition, the reading frame was predicted in each of the deletion patterns with *DMD* reading-frame checker 1.9 software.

Abbreviations: BMD, Becker muscular dystrophy; DMD, Duchenne muscular dystrophy.

instability of intron 44. We further performed flanking sequence analysis around the breakpoints in all of the 37 patients, and the results illustrated that *DMD* gene instability did not arise from a single underlying cause. We found that long motifs, nonconsensus microhomologies, low-copy repeats, palindromic sequences, and

microindels embedded around the breakpoints may predispose *DMD* to instability. Additionally, we checked all 37 breakpoint pairs in the *DMD* Open-access Variant Explorer database (www.dmd.nl), however, no identical location was found. Although the breakpoints in the hotspot regions were relatively random and few identical

TABLE 2 Microhomologies in junction reads

ID	Microhomology	Extended microhomology	Inserted sequence	M1	M2	Breakpoint in the upstream repeat region	Breakpoint in the downstream repeat region
D1	AA				Y		
D2			CTA	Y	Y		
D3						ccccaaccct gattccaaca	gtctgttgg tgtttttgt
D4	GA						
D5			CAAGTTAT			tctcaactta tcaaccaagt	atttcaaca aaatttaaaa
D6	TTA						
D7			AT				
D8	GAAATG						
D9	ACT			Y	Y		
D10	CTA						
D11				Y	Y		
D12	ACT						
D13	GA						
D14	G						
D15	CCA						
D16	CTC						
D17	TGAATA						
D18						caagcatatg aaaaaaagct	taatttaact cttctaact
D19			GTG				
D20	TTCAG				Y		
D21		CAGTA=CTATA					
D22						tttctaatac cttcagttat	aatcaactcc aaaaggaacc
D23		AAAT=GAAT					
D24	GGGT						
D25			A	Y	Y		
D26	ACC						
D27						gtcatcaatt cctgaaatac	attgcaaaa aacatttg
D28		GTTTT=ACTTT					
D29		GTTTT=ACTTT					
D30	ATT				Y		
D31						tgagaacttc tataatgttg	tagtaagtaa caaagatggc
D32	TGA			Y	Y		
D33					Y	ctagatcag tgtcctcaag	cactgcaccc cacaagttt
D34	TGC						
D35	TC			Y			
D36	CT						
D37				Y	Y		

Note: The characteristics of the flanking reads were identified, including nonconsensus microhomologies, extended microhomology, microindels, long-sequence motifs, and low-copy repeats.

Abbreviations: M1, motif A; M2, motif B; Y, yes.

breakpoints were found; in our study, one breakpoint (c.6615-1512_7660+21124del) was observed in two unrelated patients. Single-nucleotide polymorphisms (rs17338535, rs1112433, and rs17338570) in the *DMD* gene confirmed that there was no consanguinity between these two patients (Figure S5). Therefore,

to the best of our knowledge, this is the same rare breakpoint that was found in two unrelated individuals. In the present study, we observed that most of the breakpoints present in patients with DMD/BMD were not recurrent. The nonrecurrent breakpoint revealed that novel deletion variants in the dystrophin gene were

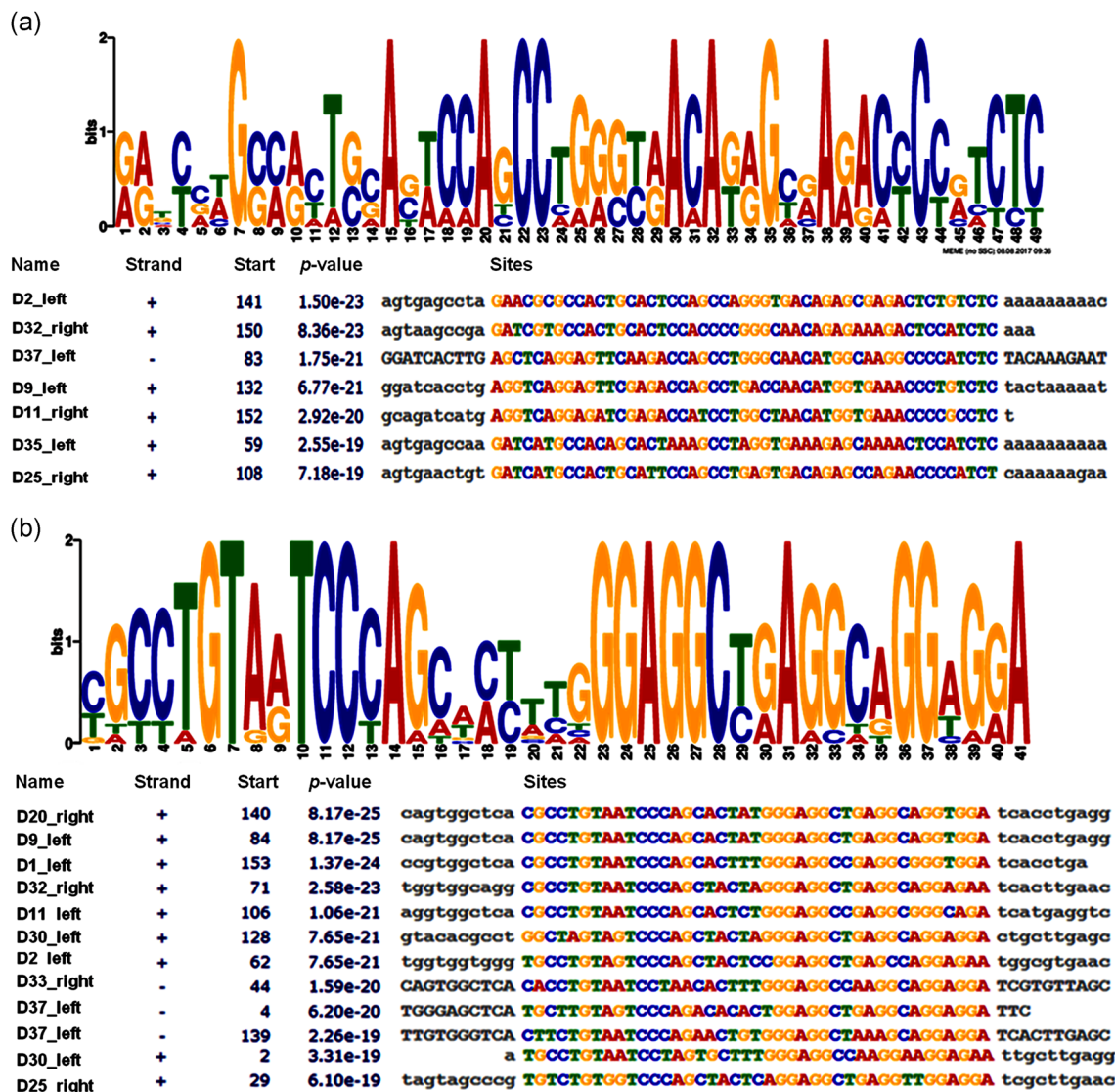


FIGURE 2 The major long-sequence motifs. (a) Motif A and the patients with the motif around the breakpoints. Left means 5-prime, and right means 3-prime of the deletion fragment. “Strand +” means the motif site was found in the sequence as it was supplied. “Strand -” means that the motif site was found in the reverse complement of the supplied sequence. “Start” means the position in the sequence where the motif site starts. “p Value” is the probability that an equal or better site would be found in a random sequence of the same length conforming to the background frequencies. “Sites” means a motif site with the 10 flanking bases on either side. (b) Motif B and the patients with the motif around the breakpoints

much more frequent than duplications or small variations (Zimowski, Pawelec, Purzycka, Szirkowicz, & Zaremba, 2017).

Most previous studies have reported the types of variants in the *DMD* gene, but only a few studies have explored the mechanisms underlying *DMD* gene instability. A study evaluating genomic instabilities in the *DMD* gene in germ cells and cancer cell lines reported that the *DMD* sequence breakpoints shared some similar features between both cell types and that the microhomologies were frequently associated with a majority of the junctions; however, the short motif was not overrepresented around the breakpoint region (Mitsui et al., 2010). Thus, to understand the potential mechanism(s) underlying large-scale variants in the *DMD* gene, we performed long-motif and microhomology analyses around the breakpoints. We

predicted that the long-motif analysis would help us to characterize the flanking sequences (200 bp) around a breakpoint. The long-motif analysis indicated the following two motifs: motif A (49 bp) was predicted to be the spliceosomal complex. A spliceosome is composed of five small nuclear RNAs (snRNAs) and a range of associated protein factors (Will & Luhrmann, 2011). It excises introns from pre-messenger RNA (mRNAs) using two sequential transesterification steps, namely branching and exon ligation (Wahl, Will, & Luhrmann, 2009). One of the most distinguishing features of the spliceosome is that it assembles stepwise on a pre-mRNA (Matera & Wang, 2014). The motif A that we found in the patient with DMD/BMD might play a similar role along with the snRNA of a spliceosome and initiate splicing. Motif B (41 bp) was predicted to be

involved in nucleotide and nucleic acid metabolic processes, which might improve DNA fragment metabolism, and play a role in *DMD* recombination. Additionally, it was predictable that the long-sequence motifs appeared to map to AluS and AluY elements, since Alu sequences are found in almost all of the known gene introns. Because we know that the Alu sequence is mainly transcribed into transfer RNA and microRNA, these RNAs might react with the template DNA locus and cause the sequence instability and even breakage. In addition, the two long motifs spread were found in ~30% of the *DMD*/*BMD* cases, which implied that long motifs might contribute to the recombination of the *DMD* gene. In previous studies on DNA double-strand breaks (DSBs) during meiotic recombination, some consensus motifs were identified (Myers, Bottolo, Freeman, McVean, & Donnelly, 2005; Myers, Freeman, Auton, Donnelly, & McVean, 2008; Sandovici et al., 2006), and the breaking mechanism was associated with the meiosis-specific topoisomerase-II-like SPO11 endonuclease (Baudat et al., 2010; Baudat, Imai, & de Massy, 2013). However, the mechanism by which DSB initiates mitotic recombination is still unclear. Consensus motifs within the breakpoint region in patients with Beckwith–Wiedemann syndrome have been reported; however, these motifs were present only in some patients (Ohtsuka et al., 2016). We speculate that large-scale deletions in the *DMD* gene could occur during both the meiotic and mitotic periods; however, the underlying mechanisms for this are complex and not clearly understood (Mitsui et al., 2010). There are at least five categories of mutational mechanisms known to initiate genomic recombination: (a) homologous recombination including nonallelic homologous recombination, gene conversion, single-strand annealing, and break-induced replication, (b) nonhomologous end joining (NHEJ), (c) microhomology-mediated replication-dependent recombination (MMRDR), (d) long interspersed element-1-mediated retrotransposition, and (e) telomere healing (Chen, Cooper, Ferec, Kehrer-Sawatzki, & Patrinos, 2010). Among these, MMRDR and NHEJ are the two main mechanisms involved in *DMD* intragenic deletions (Ishmukhametova et al., 2012; Lieber, 2008; Oshima et al., 2009). MMRDR was hypothesized to cause replication fork stalling and template switching, which could induce complex deletion and duplication rearrangement (Lee, Carvalho, & Lupski, 2007). In the present study, microhomologies or extended microhomologies were found in ~65% of *DMD* intragenic deletion patients. NHEJ explains the nonrecurring rearrangements with minimal to no junction homology. Notably, there are two mechanisms for linking DNA molecules: (a) direct ligation of ends, and (b) repair synthesis primed by terminal homologies of a few nucleotides (Roth, Porter, & Wilson, 1985). In the present study, some of the 200 flanking regions of the breakpoints were observed to be palindromic (CTTC) and contained low-copy repeat sequences (AAAA), which could promote DNA instability (Ankala et al., 2012; Marey et al., 2016; Sen et al., 2006) and then produce a direct ligation. Additionally, small-sequence insertions and deletions were frequently found (in 14% of patients with *DMD*/*BMD*), which could be explained by repair synthesis primed by terminal homologies. Studies on fragile genomic regions revealed that breakpoint distribution strongly correlated with the

length of noncoding spacers (Berthelot, Muffato, Abecassis, & Roest Crollius, 2015). In this study, we observed that all breakpoints occurred in intron regions. Intron 44 was the most frequently occurring break region. Notably, intron 44 has a longer noncoding sequence than those of the other introns in the *DMD* gene; however, the second most frequently occurring break region was in intron 50, which was much shorter than intron 44. Therefore, in this case, the criterion of the length of the noncoding spacer correlating with the breakpoint distribution was not completely satisfied in *DMD* (Figure S3a–c).

In this study, we found that although the breakpoints were present in the hotspots of the *DMD* gene, especially around introns 44 and 50, the underlying mechanisms were extremely complex. This implies that the *DMD* intragenic deletions and recombinations were associated not with a single mechanism but with a variety of factors, including long-sequence motifs, low-copy repeats, and palindromic sequences, which might promote DNA instability and induce *DMD* breaking and recombination. The breakpoints and long-sequence motifs identified in this study provide valuable new data that improve our understanding of the potential underlying mechanism. Further investigations should be performed in more *DMD*/*BMD* cases, and in-depth studies should focus on the biological function(s) of long motifs and potential coactivators, which might contribute to *DMD* gene breaking. In addition, the specific breakpoints found here might provide new insight into the development of exon-skipping therapies (van Deutekom et al., 2001) for Asian patients with *DMD*/*BMD*.

ACKNOWLEDGMENTS

The authors would like to thank the patients for participating in the *DMD* gene instability study and providing blood samples. They would also like to thank Weimin Zhang for help with sample collection. This study was supported by the Chinese Academy of Medical Sciences (CAMS), Initiative for Basic Scientific Research (Grant Numbers: 2015PT320017 and 2016RC310006).

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

DATA AVAILABILITY STATEMENT

All of the NGS data were uploaded to the Sequence Read Archive (SRA) database, and the access number is PRJNA503498.

ORCID

Chao Ling  <http://orcid.org/0000-0001-5328-4906>

REFERENCES

Ankala, A., Kohn, J. N., Hegde, A., Meka, A., Ephrem, C. L., Askree, S. H., ... Hegde, M. R. (2012). Aberrant firing of replication origins potentially

- explains intragenic nonrecurrent rearrangements within genes, including the human DMD gene. *Genome Research*, 22(1), 25–34. <https://doi.org/10.1101/gr.123463.111>
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., ... Noble, W. S. (2009). MEME suite: Tools for motif discovery and searching. *Nucleic Acids Research*, 37, W202–W208. <https://doi.org/10.1093/nar/gkp335>
- Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., ... de Massy, B. (2010). PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science*, 327(5967), 836–840. <https://doi.org/10.1126/science.1183439>
- Baudat, F., Imai, Y., & de Massy, B. (2013). Meiotic recombination in mammals: Localization and regulation. *Nature Reviews Genetics*, 14(11), 794–806. <https://doi.org/10.1038/nrg3573>
- Berthelot, C., Muffato, M., Abecassis, J., & Roest Crollius, H. (2015). The 3D organization of chromatin explains evolutionary fragile genomic regions. *Cell Reports*, 10(11), 1913–1924. <https://doi.org/10.1016/j.celrep.2015.02.046>
- Bushby, K., Finkel, R., Birnkrant, D. J., Case, L. E., Clemens, P. R., Cripe, L., ... Constantin, C. DMD Care Considerations Working Group (2010). Diagnosis and management of Duchenne muscular dystrophy, part 1: Diagnosis and pharmacological and psychosocial management. *The Lancet Neurology*, 9(1), 77–93. [https://doi.org/10.1016/S1474-4422\(09\)70271-6](https://doi.org/10.1016/S1474-4422(09)70271-6)
- Casper, A. M., Nghiem, P., Arlt, M. F., & Glover, T. W. (2002). ATR regulates fragile site stability. *Cell*, 111(6), 779–789.
- Chen, J. M., Cooper, D. N., Ferec, C., Kehrer-Sawatzki, H., & Patrinos, G. P. (2010). Genomic rearrangements in inherited disease and cancer. *Seminars in Cancer Biology*, 20(4), 222–233. <https://doi.org/10.1016/j.semcancer.2010.05.007>
- van Deutekom, J. C., Bremmer-Bout, M., Janson, A. A., Ginjaar, I. B., Baas, F., den Dunnen, J. T., & van Ommen, G. J. (2001). Antisense-induced exon skipping restores dystrophin expression in DMD patient derived muscle cells. *Human Molecular Genetics*, 10(15), 1547–1554.
- Ishmukhametova, A., Khau Van Kien, P., Mechin, D., Thorel, D., Vincent, M. C., Rivier, F., ... Tuffery-Giraud, S. (2012). Comprehensive oligonucleotide array-comparative genomic hybridization analysis: New insights into the molecular pathology of the DMD gene. *European Journal of Human Genetics*, 20(10), 1096–1100. <https://doi.org/10.1038/ejhg.2012.51>
- Lee, J. A., Carvalho, C. M., & Lupski, J. R. (2007). A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell*, 131(7), 1235–1247. <https://doi.org/10.1016/j.cell.2007.11.037>
- Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26(5), 589–595. <https://doi.org/10.1093/bioinformatics/btp698>
- Lieber, M. R. (2008). The mechanism of human nonhomologous DNA end joining. *Journal of Biological Chemistry*, 283(1), 1–5. <https://doi.org/10.1074/jbc.R700039200>
- Marey, I., Ben Yaou, R., Deburgrave, N., Vasson, A., Nectoux, J., Leturcq, F., ... Cossee, M. (2016). Non random distribution of DMD deletion breakpoints and implication of double strand breaks repair and replication error repair mechanisms. *Journal of Neuromuscular Diseases*, 3(2), 227–245. <https://doi.org/10.3233/JND-150134>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, 17(1), 10–12. <https://doi.org/10.14806/ej.17.1.200>
- Matera, A. G., & Wang, Z. (2014). A day in the life of the spliceosome. *Nature Reviews Molecular Cell Biology*, 15(2), 108–121. <https://doi.org/10.1038/nrm3742>
- Mitsui, J., Takahashi, Y., Goto, J., Tomiyama, H., Ishikawa, S., Yoshino, H., ... Tsuji, S. (2010). Mechanisms of genomic instabilities underlying two common fragile-site-associated loci, PARK2 and DMD, in germ cell and cancer cell lines. *The American Journal of Human Genetics*, 87(1), 75–89. <https://doi.org/10.1016/j.ajhg.2010.06.006>
- Myers, S., Bottolo, L., Freeman, C., McVean, G., & Donnelly, P. (2005). A fine-scale map of recombination rates and hotspots across the human genome. *Science*, 310(5746), 321–324. <https://doi.org/10.1126/science.1117196>
- Myers, S., Freeman, C., Auton, A., Donnelly, P., & McVean, G. (2008). A common sequence motif associated with recombination hot spots and genome instability in humans. *Nature Genetics*, 40(9), 1124–1129. <https://doi.org/10.1038/ng.213>
- Nobile, C., Toffolatti, L., Rizzi, F., Simionati, B., Nigro, V., Cardazzo, B., ... Danieli, G. A. (2002). Analysis of 22 deletion breakpoints in dystrophin intron 49. *Human Genetics*, 110(5), 418–421. <https://doi.org/10.1007/s00439-002-0721-7>
- Ohtsuka, Y., Higashimoto, K., Oka, T., Yatsuki, H., Jozaki, K., Maeda, T., ... Soejima, H. (2016). Identification of consensus motifs associated with mitotic recombination and clinical characteristics in patients with paternal uniparental isodisomy of chromosome 11. *Human Molecular Genetics*, 25(7), 1406–1419. <https://doi.org/10.1093/hmg/ddw023>
- Okubo, M., Goto, K., Komaki, H., Nakamura, H., Mori-Yoshimura, M., Hayashi, Y. K., ... Nishino, I. (2017). Comprehensive analysis for genetic diagnosis of Dystrophinopathies in Japan. *Orphanet Journal of Rare Diseases*, 12(1), 149. <https://doi.org/10.1186/s13023-017-0703-4>
- Onengut, S., Kavaslari, G. N., Battaloglu, E., Serdaroglu, P., Deymeer, F., Ozdemir, C., ... Tolun, A. (2000). Deletion pattern in the dystrophin gene in Turks and a comparison with Europeans and Indians. *Annals of Human Genetics*, 64, 33–40.
- Oshima, J., Magner, D. B., Lee, J. A., Breman, A. M., Schmitt, E. S., White, L. D., ... del Gaudio, D. (2009). Regional genomic instability predisposes to complex dystrophin gene rearrangements. *Human Genetics*, 126(3), 411–423. <https://doi.org/10.1007/s00439-009-0679-9>
- Ramos, E., Conde, J. G., Berríos, R. A., Pardo, S., Gomez, O., & Mas Rodriguez, M. F. (2016). Prevalence and genetic profile of Duchene and Becker muscular dystrophy in Puerto Rico. *Journal of Neuromuscular Diseases*, 3(2), 261–266. <https://doi.org/10.3233/JND-160147>
- Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29(1), 24–26. <https://doi.org/10.1038/nbt.1754>
- Roth, D. B., Porter, T. N., & Wilson, J. H. (1985). Mechanisms of nonhomologous recombination in mammalian cells. *Molecular and Cellular Biology*, 5(10), 2599–2607.
- Sandovici, I., Kassoovska-Bratinova, S., Vaughan, J. E., Stewart, R., Leppert, M., & Sapienza, C. (2006). Human imprinted chromosomal regions are historical hot-spots of recombination. *PLoS Genetics*, 2(7), e101. <https://doi.org/10.1371/journal.pgen.0020101>
- Sen, S. K., Han, K., Wang, J., Lee, J., Wang, H., Callinan, P. A., ... Batzer, M. A. (2006). Human genomic deletions mediated by recombination between Alu elements. *The American Journal of Human Genetics*, 79(1), 41–53. <https://doi.org/10.1086/504600>
- Thorvaldsdottir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2), 178–192. <https://doi.org/10.1093/bib/bbs017>
- Tuffery-Giraud, S., Beroud, C., Leturcq, F., Yaou, R. B., Hamroun, D., Michel-Calemard, L., ... Claustres, M. (2009). Genotype-phenotype analysis in 2,405 patients with a dystrophinopathy using the UMD-DMD database: A model of nationwide knowledgebase. *Human Mutation*, 30(6), 934–945. <https://doi.org/10.1002/humu.20976>
- Vengalil, S., Preethish-Kumar, V., Polavarapu, K., Mahadevappa, M., Sekar, D., Purushottam, M., ... Nalini, A. (2017). Duchenne muscular dystrophy and Becker muscular dystrophy confirmed by multiplex ligation-dependent probe amplification: Genotype-phenotype correlation in a large cohort. *Journal of Clinical Neurology*, 13(1), 91–97. <https://doi.org/10.3988/jcn.2017.13.1.91>

- Wahl, M. C., Will, C. L., & Luhrmann, R. (2009). The spliceosome: Design principles of a dynamic RNP machine. *Cell*, 136(4), 701–718. <https://doi.org/10.1016/j.cell.2009.02.009>
- Wang, J., Mullighan, C. G., Easton, J., Roberts, S., Heatley, S. L., Ma, J., ... Zhang, J. (2011). CREST maps somatic structural variation in cancer genomes with base-pair resolution. *Nature Methods*, 8(8), 652–654. <https://doi.org/10.1038/nmeth.1628>
- Will, C. L., & Luhrmann, R. (2011). Spliceosome structure and function. *Cold Spring Harbor Perspectives in Biology*, 3(7), <https://doi.org/10.1101/cshperspect.a003707>
- Zimowski, J. G., Pawelec, M., Purzycka, J. K., Szirkowiec, W., & Zaremba, J. (2017). Deletions, not duplications or small mutations, are the predominant new mutations in the dystrophin gene. *Journal of Human Genetics*, 62, 885–888. <https://doi.org/10.1038/jhg.2017.70>
- Zimowski, J. G., Massalska, D., Holding, M., Jadczyk, S., Fidzińska, E., Lusakowska, A., ... Zaremba, J. (2014). MLPA based detection of mutations in the dystrophin gene of 180 Polish families with

Duchenne/Becker muscular dystrophy. *Neurologia i Neurochirurgia Polska*, 48(6), 416–422. <https://doi.org/10.1016/j.pjnns.2014.10.004>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Ling C, Dai Y, Fang L, et al. Exonic rearrangements in DMD in Chinese Han individuals affected with Duchenne and Becker muscular dystrophies. *Human Mutation*. 2020;41:668–677. <https://doi.org/10.1002/humu.23953>