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 Wang3,7 | Xue Zhang1,8

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
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 multixonic deletion variations. Integrative genomics viewer was used for breakpoint identification (Robinson et al., 2011; Thorvaldsdottir, Robinson, & Mesirov, 2013).
### RESULTS

#### 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).

#### 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.
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 ($p < 2.55 \times 10^{-19}$) and 41 bp ($p < 6.10 \times 10^{-19}$). 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.

**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

**4 | DISCUSSION**

In the present study, we selected 896 unrelated DMD/BMD male probands with deletion/duplication detected by MLPA to characterize the multixonic 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
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

| 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                     | Intron 45-48     | Out-of-frame | DMD       |
| D7  | Male   | 3                | c.650-53656_5922+6505delinsAT                  | Intron 7-41      | Out-of-frame | DMD       |
| D8  | Male   | 3                | c.7099-10208_7661-22210del                     | Intron 48-52     | Out-of-frame | DMD       |
| D9  | Male   | 12               | c.6615-11775_7309+7415del                      | Intron 45-50     | Out-of-frame | DMD       |
| D10 | Male   | 8                | c.7099-12971_8027+3726del                      | Intron 48-54     | Out-of-frame | DMD       |
| D11 | Male   | 10               | c.7098+12114_8027+7417del                      | Intron 48-54     | Out-of-frame | DMD       |
| D12 | Male   | 3                | c.7309+5104_7542+17062del                      | Intron 50-51     | Out-of-frame | DMD       |
| D13 | Male   | 3                | c.6614+5702_8217+43627del                      | Intron 45-55     | Out-of-frame | DMD       |
| D14 | Male   | 3                | c.6912+18064_8027+2675del                      | Intron 47-54     | Out-of-frame | DMD       |
| D15 | Male   | 3                | c.7098+7776_7660+1050del                      | Intron 48-52     | Out-of-frame | DMD       |
| D16 | Male   | 3                | c.7309+15311_7872+6382del                      | Intron 50-53     | Out-of-frame | DMD       |
| D17 | Male   | 15               | c.6438+102544_6913-15866del                    | Intron 44-47     | In-frame     | BMD       |
| D18 | Male   | 7                | c.7309+14745_8028-7983del                      | Intron 50-54     | Out-of-frame | DMD       |
| D19 | Male   | 11               | c.7099-10048_8028-4789delinsCAC               | Intron 48-54     | Out-of-frame | DMD       |
| D20 | Male   | 2                | c.7309+4829_8027+1051del                      | Intron 50-54     | Out-of-frame | DMD       |
| D21 | Male   | 9                | c.6912+3308_8027+6905del                       | Intron 47-54     | Out-of-frame | DMD       |
| D22 | Male   | 11               | c.7201-7674_7310-20273del                     | Intron 49-50     | Out-of-frame | DMD       |
| D23 | Male   | 16               | c.6439-28736_6913-10868del                     | Intron 44-47     | In-frame     | BMD       |
| D24 | Male   | 5                | c.7310-11092_8218-59513del                     | Intron 50-55     | Out-of-frame | DMD       |
| D25 | Male   | 3                | c.6614+13221_7310-2598delinsT                 | Intron 45-50     | Out-of-frame | DMD       |
| D26 | Male   | 7                | c.6913-6367_7661-17902del                     | Intron 47-52     | Out-of-frame | DMD       |
| D27 | Male   | 4                | c.7099+1406_7660+5892del                      | Intron 48-52     | Out-of-frame | DMD       |
| D28 | Male   | 4                | c.6615-1512_7660+21124del                     | Intron 45-52     | Out-of-frame | DMD       |
| D29 | Male   | 8                | c.6615-1512_7660+21124del                     | Intron 45-52     | Out-of-frame | DMD       |
| D30 | Male   | 3                | c.6439-49382_7872+308del                      | Intron 44-53     | In-frame     | BMD       |
| D31 | Male   | 3                | c.6913-9844_7310-16920del                     | Intron 47-50     | Out-of-frame | DMD       |
| D32 | Male   | 8                | c.6615-419_7200+5420del                       | Intron 45-49     | Out-of-frame | DMD       |
| D33 | Male   | 7                | c.7543-14415_8027+1540del                     | Intron 51-54     | Out-of-frame | DMD       |
| D34 | Male   | 2                | c.6439-40173_7661-320del                      | Intron 44-52     | Out-of-frame | DMD       |
| D35 | Male   | 3                | c.6439-97179_7200+8061del                     | Intron 44-49     | In-frame     | BMD       |
| D36 | Male   | 2                | c.6439-48962_7310-2533del                     | Intron 44-50     | Out-of-frame | DMD       |
| D37 | Male   | 4                | c.6439-16684_8217+15910del                    | Intron 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.
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

| 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      | GA                              |                      |                   |    |    | ccccaaccct|gattcacaaca | gcctgggt|gctttttgt |
| D4      | CAAGTTAT                        |                      |                   |    |    | tctcaactta|caaccaagt | atttcaaca|aaat taaa |
| D5      | TTA                             |                      |                   |    |    |                                     |                                     |
| D6      | AT                              |                      |                   |    |    |                                     |                                     |
| D7      | GAAATG                          |                      |                   |    |    |                                     |                                     |
| D8      | ACT                             |                      |                   |    |    |                                     |                                     |
| D9      | CTA                             |                      |                   |    |    |                                     |                                     |
| D10     | ACT                             |                      |                   |    |    |                                     |                                     |
| D11     | ACT                             |                      |                   |    |    |                                     |                                     |
| D12     | GA                              |                      |                   |    |    |                                     |                                     |
| D13     | G                               |                      |                   |    |    |                                     |                                     |
| D14     | CCA                             |                      |                   |    |    |                                     |                                     |
| D15     | CTC                             |                      |                   |    |    |                                     |                                     |
| D16     | TGAATA                          |                      |                   |    |    |                                     |                                     |
| D17     | GTG                             |                      |                   |    |    |                                     |                                     |
| D18     | TTTG                            |                      |                   |    |    |                                      |                                     |
| D19     | TTTT                            |                      |                   |    |    |                                      |                                     |
| D20     | TTCAG                           |                      |                   |    |    |                                      |                                     |
| D21     | CAGTA==CTATA                    |                      |                   |    |    |                                      |                                     |
| D22     | TTTCAACTTT                      |                      |                   |    |    |                                      |                                     |
| D23     | AAAT==GAAT                      |                      |                   |    |    |                                      |                                     |
| D24     | GGTT                            |                      |                   |    |    |                                      |                                     |
| D25     | ACC                             |                      |                   |    |    |                                      |                                     |
| D26     | ACC                             |                      |                   |    |    |                                      |                                     |
| D27     | GTTTT==ACTTT                    |                      |                   |    |    |                                      |                                     |
| D28     | GTTTT==ACTTT                    |                      |                   |    |    |                                      |                                     |
| D29     | GTTTT==ACTTT                    |                      |                   |    |    |                                      |                                     |
| D30     | ATT                             |                      |                   |    |    |                                      |                                     |
| D31     | TGAGACTTT                      |                      |                   |    |    |                                      |                                     |
| D32     | TGA                             |                      |                   |    |    |                                      |                                     |
| D33     | TGC                             |                      |                   |    |    |                                      |                                     |
| D34     | TC                              |                      |                   |    |    |                                      |                                     |
| D35     | CT                              |                      |                   |    |    |                                      |                                     |
| D36     | CT                              |                      |                   |    |    |                                      |                                     |
| D37     | CT                              |                      |                   |    |    |                                      |                                     |

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.
much more frequent than duplications or small variations (Zimowski, Pawelec, Purzycka, Szirkowiec, & 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 long-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...
**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