Journal of Human Genetics (2017) 62, 1057–1063

www.nature.com/jhg

ORIGINAL ARTICLE

Precise mapping of 17 deletion breakpoints within the central hotspot deletion region (introns 50 and 51) of the *DMD* gene

Gabriella Esposito[1,2], Maria Roberta Tremolaterra[1], Evelina Marsocci[2,4], Igor CM Tandurella[2,5], Tiziana Fioretti[3], Maria Savarese[2] and Antonella Carsana[1,2]

Exon deletions in the human *DMD* gene, which encodes the dystrophin protein, are the molecular defect in 50–70% of cases of Duchenne/Becker muscular dystrophies. Deletions are preferentially clustered in the 5′ (exons 2–20) and the central (exons 45–53) region of *DMD*, likely because local DNA structure predisposes to specific breakage or recombination events. Notably, innovative therapeutic strategies may rescue dystrophin function by homology-based specific targeting of sequences within the central *DMD* hot spot deletion region. To further study molecular mechanisms that generate such frequent genome variations and to identify residual intronic sequences, we sequenced 17 deletion breakpoints within introns 50 and 51 of *DMD* and analyzed the surrounding genomic architecture. There was no breakpoint clustering within the introns nor extensive homology between sequences adjacent to each junction. However, at or near the breakpoint, we found microhomology, short tandem repeats, interspersed repeat elements and short sequence stretches that predispose to DNA deletion or bending. Identification of such structural elements contributes to elucidate general mechanisms generating deletion within the *DMD* gene. Moreover, precise mapping of deletion breakpoints and localization of repeated elements are of interest, because residual intronic sequences may be targeted by therapeutic strategies based on genome editing correction.
Journal of Human Genetics (2017) **62,** 1057–1063; doi:10.1038/jhg.2017.84; published online 7 September 2017

INTRODUCTION

Duchenne muscular dystrophy (DMD, OMIM #310200) and Becker muscular dystrophy (BMD, OMIM #300376) are X-linked recessive neuromuscular disorders characterized by progressive muscle wasting and weakness due to degeneration of skeletal, smooth and cardiac muscles. The incidence of DMD and BMD is ~1 in 3500 and 1 in 20 000–30 000 live male births, respectively.[1] DMD occurs earlier and is more severe than BMD. Patients affected by BMD have a broad spectrum of clinical symptoms ranging from mild to severe. No effective treatments exist for the severe forms of these disorders, both caused by mutations in the *DMD* gene (OMIM *300377).

The *DMD* gene, located at Xp21 locus, spans more than 2 Mb and contains 79 relatively small exons (range 32–269 bp), whereas intron (intervening sequence, IVS) size ranges from 107 to 248 401 bp. It has at least seven different tissue-specific promoters and two polyadenylation sites; moreover, differential processing gives rise to a variety of *DMD* transcripts encoding a set of protein isoforms.[1] The protein translated from the larger transcript is the dystrophin, an important cytoskeletal protein that links the cytoskeleton of muscle fibre to the underlying basal lamina, in the skeletal muscle. Alteration or loss of dystrophin forces excess calcium into the muscle cell membrane, which causes mitochondria to undergo a 'permeability transition', that is, the regulated formation of the permeability transition pore complex or mitochondrial megachannel, which spans the outer and inner mitochondrial membranes, leading to an irreversible loss of matrix and intermembrane contents and swelling of the mitochondria;[2] thus, the affected skeletal muscle will undergo mitochondrial dysfunction, amplification of stress-induced reactive-oxygen species production, sarcolemma damages and cell death. These cellular alterations cause increased serum levels of creatine kinase and sometimes hypertransaminasemia, also in asymptomatic patients, already in early childhood.[3,4] The disease progression is characterized by widespread necrosis of muscle fibres that are ultimately replaced with adipose and connective tissue.

Exon deletions in the human *DMD* gene are the most common pathogenic variants (50–70%) in DMD/BMD; 30% of patients has point mutations and about 5% exon duplications.[5] Gene variants that shift the *DMD* open-reading frame in the spliced mRNA or that lead to stop codons result in truncated, non-functional dystrophin and usually gives rise to the DMD phenotype. Alternatively, genetic

[1]Department of Molecular Medicine and Medical Biotechnology, University of Naples Federico II, Naples, Italy; [2]CEINGE Advanced Biotechnology, Naples, Italy and [3]IRCCS SDN, Naples, Italy
[4]Current address: Laboratorio Athena, Via Ilaria Alpi, 62, 47522 Cesena, Italy.
[5]Current address: Ospedale Oncologico di Riferimento Regionale 'Armando Businco', 09121 Cagliari, Italy.
Correspondence: Professor A Carsana, Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Università di Napoli Federico II, Via San Pansini 5, 80131 Naples, Italy.
E-mail: carsana@unina.it
Received 3 May 2017; revised 6 July 2017; accepted 24 July 2017; published online 7 September 2017

Table 1 Primer sequences to amplify STS in IVS 50 and IVS 51, and amplicon regions (gene reference NG_012232.1)

STS	*Forward primer (5′–3′)*	*Reverse primer (5′–3′)*	*Amplicon*
IVS50			
A	GCAACTATGAAGTGATGACTGG	AGTTGTCCTTCCACTCTCAGGC	1 524 687–1 525 077
B	ATTCTGATGGTCTTGCTTGG	CTTAGCAGAAGTAACCTACACG	1 528 541–1 528 813
C	CCTGTGTGAGTTCAGACTTGCC	CAAGGTAATGCTTATGGACC	1 541 329–1 541 627
D	ATCGGTGAAGCATAACTGAAGG	ACTATTATTGCGGAGATGCC	1 550 125–1 550 336
E	GAAGTGAAGGAGGCAGTATTGG	TGTCCTCACTACTCCTACTTGC	1 557 712–1 558 104
F	TAGCATAGTTCCATAGCGGTGG	AATGACAACATCTCTGCTGG	1 562 217–1 562 464
G	ATAGATGTTCTTCAGGTGGTGC	GAAGGAGGAAGGAATCTATTGG	1 567 642–1 567 900
H	TCACTGAATCTACACAACTGCC	TGTAGAGTAAGTCAGCCTATGG	1 569 432–1 569 695
IVS51			
I	TACTGTGAGGTAGATAGAGAGG	TAGCATCCATTCTCTCTCCTGG	1 573 819–1 574 027
L	AGCAACTTCATTGAGGCTAAGC	GATGAAGTCTCACTCTGTTGCC	1 580 147–158 043
M	TGTATGGAGGTTCTTCTTGC	CTGAACACCTCTCAACAGAAGG	1 585 869–1 586 214
N	CTTATTACTTCCTTCTGCCTGC	GGATGTGAATGAATCAGTAGGC	1 591 236–1 591 493
O	CCATCAGAGTGAACAGGCAACC	GCCACACTGACTTCCACAATGG	1 594 498–1 594 875
P	GTTGCTGTTCCTCTATGGCACC	GCATAACACATCACAGTAGG	1 599 487–1 599 508
Q	TTGTGGCTGAGTAGTATTCC	GCACAACCACTATGGAGAACCG	1 605 863–1 606 121
S	AAGAAGTTACCGATGGAGGTGC	TCACTGAGCATCTTCTTCTGGC	1 610 659–1 610 680

Abbreviation: STS, sequences tagged site.

variants that do not affect the open-reading frame, thus predicting the production of a semi-functional protein, have been associated to the less severe BMD phenotype.[6] This reading-frame hypothesis, which arises from analysis of genomic DNA, applies to about 90% of DMD/BMD cases and is the basic principle underlying the most recent molecular strategies that aim to definitively restore the dystrophin reading frame in muscle cells of severely affected DMD patients.[7]

Notably, the evidence that intragenic deletions preferentially cluster in two areas of the *DMD* gene, that is, the 5′ (exons 2–20) and the central (exons 45–53) regions, suggests that some features of the DNA or of the chromatin structure within intronic sequences may predispose the *DMD* gene to specific breakage or recombination events. However, precise mapping of intronic deletion breakpoints revealed no clustering of breakpoints and no extensive homology between adjacent sequences to each breakpoint.[8–19] Therefore, the molecular mechanisms generating these copy number variations remain poorly understood, although models based on locus-specific genomic architecture have been proposed.[20] Moreover, because a relatively high frequency of rearrangements affects central exons of *DMD*, precise mapping of the breakpoint junctions located within this deletion rich region could drive a more precise targeting in a potential genome editing-based therapy.

Here we report mapping and sequence analysis of 17 breakpoints located in IVSs 50 and 51 of the *DMD* gene and the DNA sequence signatures at the junctions of each deletion. IVSs 50 and 51 are located within the central hotspot region that border the sequences encoding the hinge III segment (exons 50 and 51) of the dystrophin protein whose absence/presence has been correlated with the severity of the BMD phenotype.[21]

MATERIALS AND METHODS

We analyzed 17 unrelated DMD/BMD (DMD, $n=10$, BMD, $n=7$) patients carrying deletions starting or ending within IVSs 50 and 51 whose DNA samples were still available after molecular diagnosis. Exon deletions were detected by using quantitative fluorescence multiplex-PCR and multiple ligation probe amplification as described.[5,22,23] Informed consent was obtained for each patient according to the procedure established by the local Bioethics Institutional Committee.

We amplified eight sequences tagged sites (Table 1, 200–300 bp), distributed at contiguous intervals of roughly similar size, within IVS 50 and IVS 51 of the *DMD* gene (GenBank NG_012232.1), in 17 DMD/BMD male patients with deletions starting or ending in these two IVSs. Moreover, we amplified *DMD*-specific sequences in IVS 44 (248 401 bp), 47 (54 222 bp) and 48 (38 368 bp) as described elsewhere.[12,24] The absence/presence of these sequences in patients was verified by 1% agarose gel electrophoresis, to define the intronic interval where each breakpoint lies. Genomic regions spanning the deletion breakpoints were amplified by long-PCR (Expand High Fidelity PCR System, Roche Molecular Systems Inc, Pleasanton, CA, USA), cloned in the pGEM-T (Promega, Fitchburg, WI, USA) plasmid and sequenced by primer walking on the ABI-Prism 377 Applied Biosystems Genetic analyzer (Applied Biosystems Inc, Foster City, CA, USA). Structural features of intronic sequences were characterized by searching for inverted and tandem repeats using the EMBOSS software analysis package (program Palindrome, http://www.bioinformatics.nl/cgi-bin/emboss/palindrome and Etandem http://emboss.bioinformatics.nl/cgi-bin/emboss/etandem, respectively), and for repetitive elements using the RepeatMasker program (www.repeatmasker.org).

RESULTS

We mapped deletion breakpoints in 17 DMD/BMD patients that carry deletions starting or ending within IVSs 50 (45 782 bp long) and 51 (44 211 bp long) of the *DMD* gene (Figures 1 and 2). First, we used sequences tagged sites walking to roughly map residual intronic sequences in IVS 44, 47, 48, 50 and 51. Then, we amplified and sequenced the DNA fragments spanning all the deletion junctions (Figure 3, junction 1–17). Therefore, we precisely determined the size of each deletion (ranging from 2454 to 419 729 bp) and residual sequences of each IVS (Table 2). Sequencing allowed us to localize, at nucleotide level, 2 deletion startpoints in IVS 50, and 8 and 9 endpoints in IVSs 50 and 51, respectively. In these IVSs, breakpoints appear to be evenly distributed without any significant clustering also when compared with data reported by other authors.[16–19,25] Interestingly, patients 76/343 and 171/676 (J7 and J8, respectively, Table 2), both carrying the exon 48–50 deletion, have probable (end of sequence homology) startpoints in IVS 47 and endpoints in IVS 50 located at a

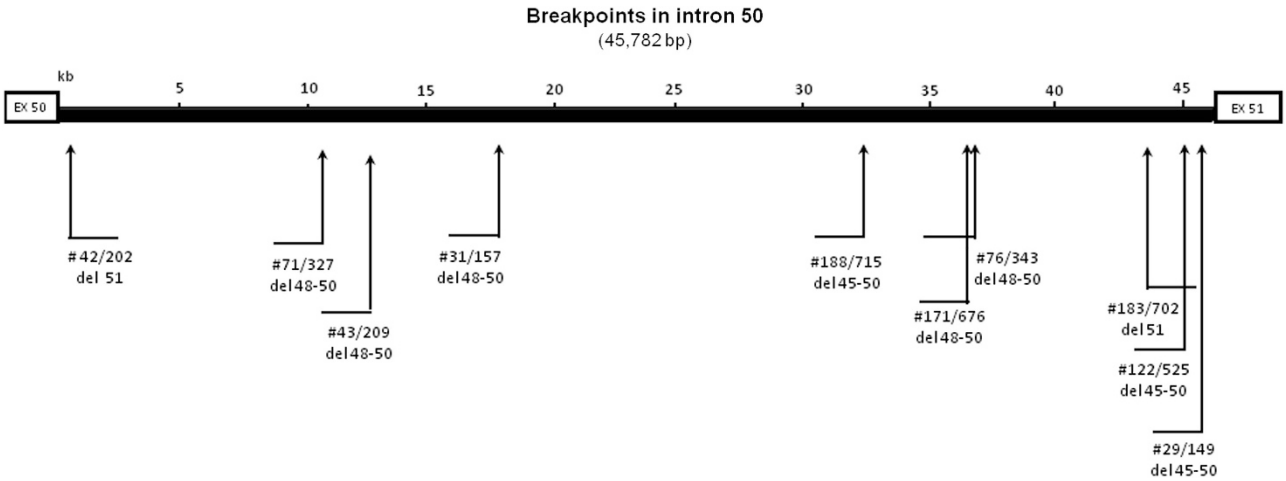

Figure 1 Localization of deletion breakpoints in IVS 50. The sample identification numbers and the *DMD* exon deletions are indicated; ↑ deletion startpoint; ↑ deletion endpoint.

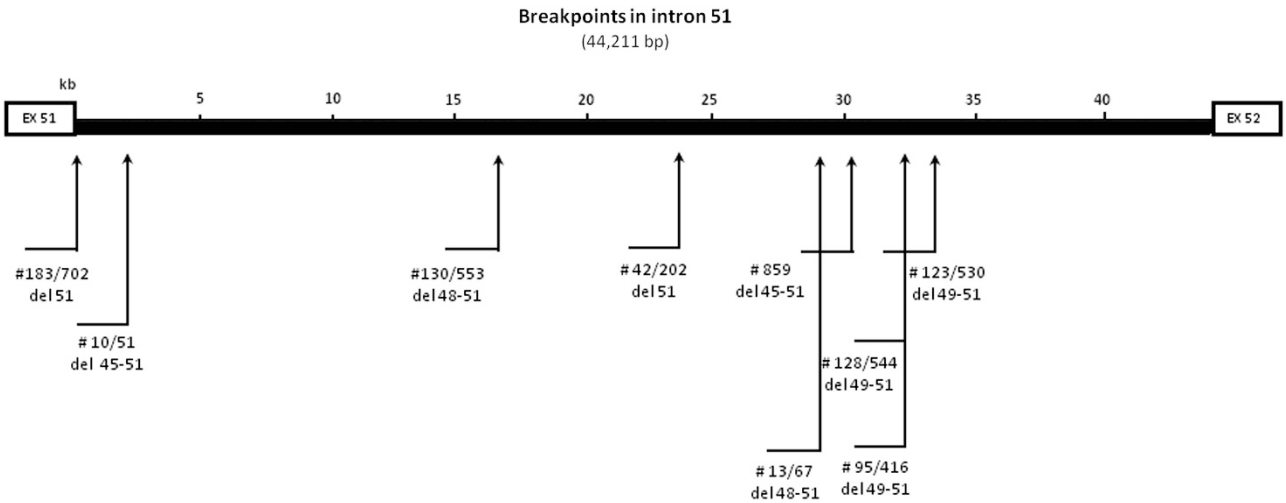

Figure 2 Localization of deletion breakpoints in IVS 51. The sample identification numbers and the *DMD* exon deletions are indicated; ↑ deletion startpoint; ↑ deletion endpoint.

distance of 12 and 14 nt, respectively (Figure 3, Table 2). Moreover, patients 128/544 and 95/416 (J16 and J17, respectively, Table 2), both carrying the deletion of exons 49–51, share the same startpoint in IVS 48 and have probable endpoints in IVS 51 located at distance of only 3 nt (Figure 3 and Table 2). Microhomologies of 1–5 bp are present in 58.8% (10/17) of the breakpoint junctions (Figure 3 and Table 2) and small insertions (1–22 bp) have been found in 5 cases (J3, J7, J12, J13 and J15, Figure 3 and Table 2). Nucleotides from 4 to 16 of the 20 bp sequence inserted within J12 (patient 859, deletion of exons 45–51) are identical to a sequence of the *ZFX* gene (OMIM *314980), at chromosome Xp22.11 (Chromosome X: 24,149,173–24,216,255), about 7 Mb away from the *DMD* gene; the longest (22 bp) inserted sequence is present in J13 (patient 123/530, deletion of exons 45–51) and is identical to a sequence located at IVS 44 of the *DMD* gene, about 1300 nt downstream the deletion startpoint. Moreover, deletions of 1 (J1, J3, J4 and J15), 3 (J5), 11 (J8) and 56 (J15) nt have been observed within 30 nt from the probable breakpoints (Figure 3 and Table 2). Genomic rearrangements can give rise to microdeletions, as the 11 and 56 nt deletions we detected in J8 and J15 breakpoints, which abolish a palindromic sequence (Figure 3) and lie in the 3′ of a LINE/L1 repeat sequence (Table 2), respectively. Sequence features that might facilitate DNA breakage through exposure of single-stranded DNA, such as inverted repeated sequences with a stem of 6 or more base pairs, are present near 12/17 breakpoints (J1, J2, J4, J7, J8, J9, J10, J11, J14, J15, J16 and J17) and short (6–8 nt) tandem repeats are present near the breakpoints in junctions J5, J6, J7, J8, J9 and J13 (Figure 3). The short deletion consensus sequence TGRRKM[26] is present in proximity of 10 deletion breakpoints (J2, J5, J7, J8, J10, J11, J12, J14, J15 and J17, Figure 3) and the immunoglobulin heavy chain class switch repeat TGAGC,[27] which has been found over-represented in close proximity to many deletion breakpoints,[28] is present in 3 deletion breakpoints (J7, J8 and J12, Figure 3). The TTTAAA sequence, which is known to be able to curve the DNA molecule,[29] has been found at or in close proximity to junctions J4, J10 and J14 (Figure 3). In particular, three TTTAAA sequences are present in junction J4: two flank the breakpoint site at IVS47, and one is in IVS50 (Figure 3). Analysis of IVS 50 and 51 sequences of the *DMD* gene revealed a number of interspersed repeats, such as LINE, LTR and SINE, as reported in Table 2; ~82% (14/17) of sequenced breakpoints are flanked by at least one kind of repeat element (Table 2). However, no extensive homology exists in these cases.

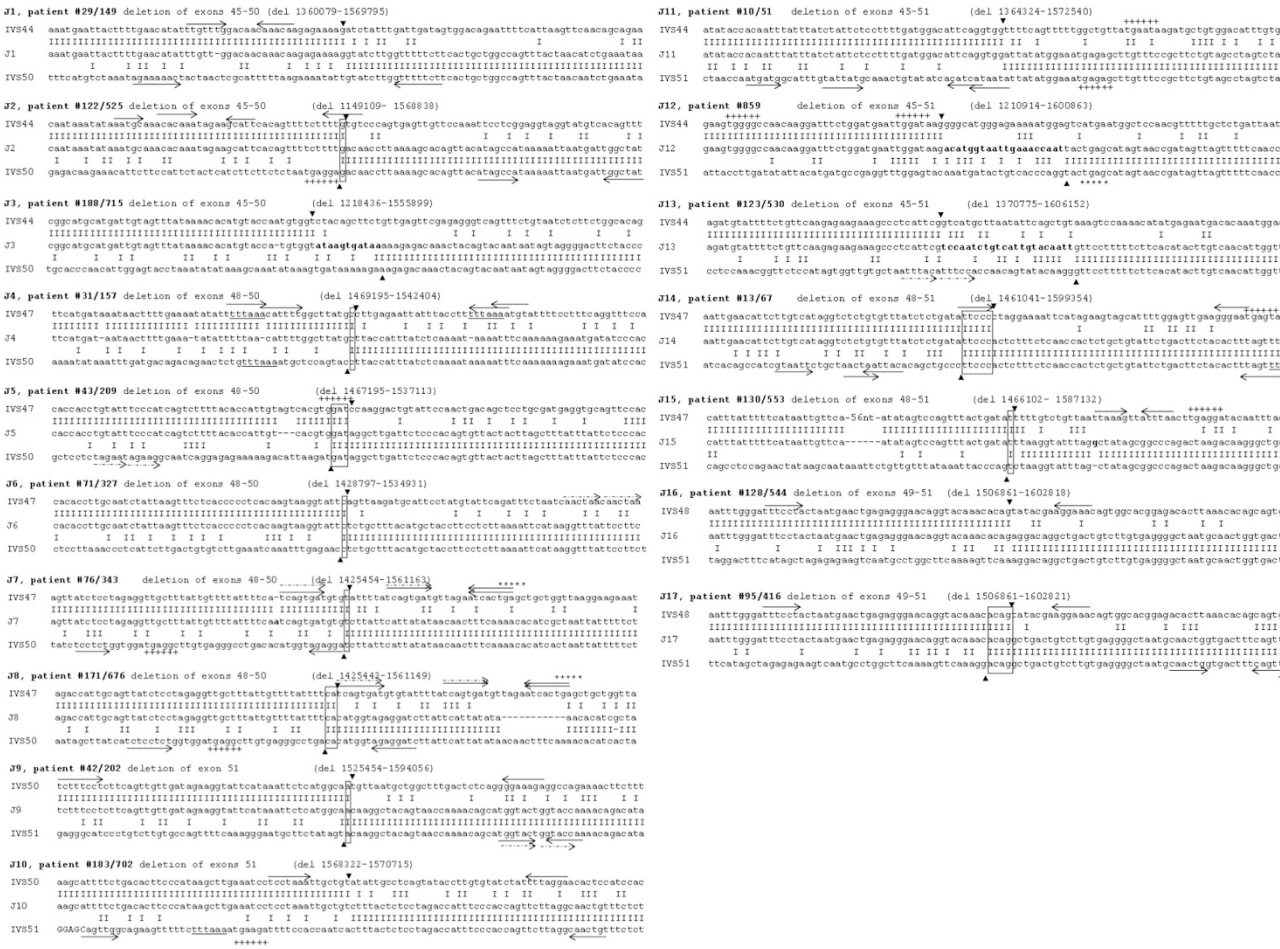

Figure 3 DNA sequences spanning the 17 deletion junctions (J1–J17), with the corresponding normal intervening sequences (IVS) regions. Only the 5′–3′ strands are shown. The numeration was on the dystrophin genomic reference sequence NG_012232.1. Capital letters indicate exonic sequence (J10). Microhomology regions across the deletion junction are in boxes. Filled arrowheads (▼) show the probable deletion breakpoints (end of sequence homology). New nucleotides inserted are in bold. The sequence TTTAAA is underlined. Arrows and dotted arrows indicate palindromic sequences with a stem of 6 or more base pairs and short tandem repeats, respectively; $^{+++++}$ indicates the short deletion consensus, TGRRKM;[26] **** indicates the immunoglobulin class switch repeat TGAGC.[27] 56 nt: 56 nt deletion in J15 (IVS 47, nt 51010–51065).

DISCUSSION

In this work we report molecular and structural characterization of 17 deletion breakpoints that start or end at IVS 50 and 51 of the *DMD* gene. Several mechanisms have been proposed to explain genomic rearrangements, such as non-allelic homologous recombination, non-homologous end joining (NHEJ), microhomology-mediated end joining (MMEJ) involving short (2–20 bp) stretches of sequence and L1 retrotransposition.[20] In particular, NHEJ is one of the main mechanisms for repairing double-stranded DNA breaks by ligation of two DNA ends without the need for a homologous template; the two DNA ends then associate in a manner that tolerates nucleotide loss or addition at the junction site. NHEJ typically utilizes short homologous DNA sequences to guide repair. Some features of the DNA breakpoint sequence we analysed could contribute to generate *DMD* gene rearrangements. Indeed, microhomologies of 1–5 bp, typical of NHEJ and MMEJ events, are present in 58.8% of our deletion junctions, comparable to the previous findings for the *DMD* gene.[17,30,31] Furthermore, the local DNA sequence environment may be involved in promoting gene deletions. Human sequences involved in deletion events often contain the short deletion consensus sequence, TGRRKM, generally associated with palindromic sequences at junctions.[26] We found that several deletion consensus sequences, palindromic sequences and short tandem repeats flank the junctions we analyzed. These sequences may promote DNA instability by facilitating the formation of secondary structure intermediates (hairpin loop structure in a single strand of DNA), which predispose DNA to breakage and intragenic recombination. Moreover, five TTTAAA sequences, known to be able to induce DNA bending, have been found at or in close proximity of junctions J4, J10 and J14 (Figure 3).[29] Given the expected frequency of this sequence (1/1420 bp) in the human genome,[32] the probability that 5 such sequences occurred by chance near the breakpoints is extremely low, suggesting that curved DNA may be involved in one of the mechanisms of the *DMD* gene deletions. Furthermore, RepeatMasker analysis revealed that 82.3% of breakpoint junctions aligned with one or more repeat elements, consistent with previous studies on *DMD* gene rearrangements, but much higher than the average frequency (35.6%) of repetitive sequences in the *DMD* gene.[6,12] Therefore, these highly homologous sequences (such as LINEs or SINE) might be involved in *DMD* gene rearrangements. Interestingly, no such repeated element was identified around the almost equal junctions J7 and J8 (Figure 3), which obviously share various DNA motifs (inverted and palindromic

Table 2 Breakpoints and DNA sequence signatures at the junctions of 17 deletions in the *DMD* gene

Junction	Sample ID	Deleted exons	BP location [a]	Deletion size (bp)	Intronic residual sequences (bp) 5′-IVS	3′-IVS	MH (bp)	Insertions and deletions	RepeatMasker analysis (±150 bp) Proximal BP	Distal BP
J1	29/149	45–50	1 360 079–1 569 795	209 717	IVS 44 237 705	IVS 50 622		Deletion:1 nt (t) in IVS 44, 21 nt upstream the probable breakpoint		
J2	122/525	45–50	1 149 109–1 568 838	419 729	IVS 44 21 415	IVS 50 1 579	g		A-rich: low_complexity LINE:L2	
J3	188/71	45–50	1 218 436–1 555 899	337 463	IVS 44 90 742	IVS 50 14 518		Deletion: 1 nt (a) in IVS 44, 6 nt upstream the probable breakpoint. Insertion: 11 nt (ataagtgataa)		LINE:L1
J4	31/157	48–50	1 469 195–1 542 404	73 209	IVS 47 54 181	IVS 50 28 013	c	Deletions: 1 nt (a) in IVS 44, 14 nt upstream the probable breakpoint and 1 nt (a) in IVS 50 20 nt downstream the probable breakpoint		LINE:L2
J5	43/209	48–50	1 467 195–1 537 113	69 916	IVS 47 52 181	IVS 50 33 304	gat	Deletion: 3 nt (agt) in IVS47, 9 nt upstream the probable breakpoint	LTR:ERV1	(TTTTC)n
J6	71/327	48–50	1 428 796–1 534 931	106 135	IVS 47 13 782	IVS 50 35 486	c		LTR:ERVL-MaLR	
J7	76/343	48–50	1 425 454–1 561 163	135 709	IVS 47 10440	IVS 50 9 254	t	Insertion: 1 nt (a) in IVS 47, 11 nt upstream the probable breakpoint		
J8	171/676	48–50	1 425 443–1 561 149	135 705	IVS 47 10 429	IVS 50 9 268	ca	Deletion: 11 nt (acaactttcaa) in IVS 50, 30 nt downstream the probable breakpoint		
J9	42/202	51	1 525 453–1 594 056	68 603	IVS 50 817	IVS 51 20 805	a		(TTTTTG)n	LINE:L1
J10	183/702	51	1 568 322–1 570 775	2454	IVS 50 43 686	IVS 51 44 086			LINE:CR1	(A)n Simple repeat
J11	10/51	45–51	1 364 326–1 572 540	208 215	IVS 44 236 632	IVS 51 42 321			LINE:L1	SINE/MIR
J12	859	45–51	1 210 916–1 600 863	389 948	IVS 44 83 222	IVS 51 13 998		Insertion: 20 nt (acatggtaattgaaaccaat)	LINE/L2	LINE:L1
J13	123/530	45–51	1 370 775–1 606 152	235 378	IVS 44 243 081	IVS 51 8 709		Insertion: 20 nt (tccaatctgtcattgtacaatt)	LINE:L1	LINE:L1
J14	13/67	48–51	1 461 041–1 599 354	138 309	IVS 47 46 027	IVS 51 15 507	ttccc		LINE:L1	
J15	130/553	48–51	1 466 104–1 587 132	121 028	IVS 47 51 090	IVS 51 27 729	t	Deletion: 56 bp in IVS47 (nt 1466024 to 1466078), 23 nt upstream the probable breakpoint. Insertion: 1 nt (g) in IVS 51 14 nt downstream the probable breakpoint	LINE: L1	LTR:MaLR LTR:ERVL
J16	128/544	49–51	1 506 861–1 602 818	95 958	IVS 48 37 439	IVS 51 12 043				DNA/MER2_type
J17	95/416	49–51	1 506 861–1 602 821	95 957	IVS 49 37 439	IVS 51 12 040	acag			DNA/MER2_type

Abbreviations: BP, breakpoint; MH, microhomology.
[a]The numeration was on the *DMD* genomic reference sequence NG_012232.1.

repeats). This observation supports the concept that genetic instability may depend on short DNA motifs that mark the genomic architecture of such regions.

Although the limited number of samples we analyzed may reduce the significance of our results, the sequence signature analysis we performed at level of the breakpoint junctions of 17 *DMD* gene deletions is helpful in explaining why rearrangements preferentially occur within particular *DMD* gene regions. Moreover, it enlarges the number of sequenced deletion breakpoints in IVSs 50 and 51, which are located within the hotspot deletion region and border the sequences encoding the hinge III segment (exons 50 and 51) of the dystrophin. The absence/presence of the hinge III has been correlated with the severity of the BMD phenotype; in fact, the loss of this region correlates with a slower disease progression.[21] Therefore, removing DNA regions that encode this protein domain or restoring the reading frame of out-of-frame deletions by mutation-specific gene therapy, including exon skipping or suppression of nonsense mutations during translation, may be considered a strategy to treat severe DMD forms.[7,33] Efficient genomic correction using CRISPR–Cas9 system is a promising method to permanently correct *DMD* mutations in induced pluripotent stem cells of DMD patients at genomic level.[33] Indeed, genome editing by CRISPR–Cas9 is able to remove internal, but unessential, regions of the mutated gene to restore the reading frame.[6] Driver of the editing is the Cas9 protein that forms a complex with the 3′ end of a single guide RNA (sgRNA) molecule. The protein-RNA pair recognizes its genomic target by complementary base pairing between the 5′-end of the sgRNA sequence and a predefined 20 bp DNA sequence, known as protospacer, which must be identified within the target genomic region. Alternatively, co-introducing in the cell Cas9/sgRNA and a donor DNA template can induce a homologous recombination event that allows to insert a specific DNA element at a target locus, based on the homology sequence of the 5′- and 3′-arms of the donor template.[33]

Notably, reading frame could be efficiently restored by removing sequences encoding exon 51 in our DMD patients with the out-of-frame deletions of exons 45–50 and 48–50 (J1–J8, Figure 3 and Table 2). Moreover, full-length dystrophin could be rescued by knocking-in exon 51 in the two patients with the isolate deletion of exon 51 (J9 and J10, Table 2). In such cases, potential targets of genome editing lie within residual sequences of IVS50, exon 51 and IVS51 of *DMD* and should spare repeated elements. Therefore, exact breakpoint mapping and information about the genomic architecture surrounding breakpoint junctions could drive a more precise and specific targeting of the residual intronic sequences within the central *DMD* deletion rich region in a potential genome editing-based therapy.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work has been supported by Ministry of Health, Italy, grant RF-2011-02349269 to GE and by Regione Campania (Convenzione CEINGE-Regione Campania, G.R: 27/12/2007).

1 Muntoni, F., Torelli, S. & Ferlini, A. Dystrophin and mutations: one gene, several proteins, multiple phenotypes. *Lancet Neurol.* **2**, 731–740 (2003).
2 Zamzami, N. & Kroemer, G. The mitochondrion in apoptosis: how Pandora's box opens. *Nat. Rev. Mol. Cell. Biol.* **2**, 67–71 (2001).
3 Paolella, G., Pisano, P., Albano, R., Cannaviello, L., Mauro, C., Esposito, G. *et al.* Fatty liver disease and hypertransaminasemia hiding the association of clinically silent Duchenne muscular dystrophy and hereditary fructose intolerance. *Ital. J. Pediatr.* **38**, 64 (2012).
4 Veropalumbo, C., Del Giudice, E., Esposito, G., Maddaluno, S., Ruggiero, L. & Vajro, P. Aminotransferases and muscular diseases: a disregarded lesson. Case reports and review of the literature. *J. Paediatr. Child Health* **48**, 886–890 (2012).
5 Carsana, A., Frisso, G., Intrieri, M., Tremolaterra, M. R., Savarese, G., Scapagnini, G. *et al.* A 15-year molecular analysis of DMD/BMD: genetic features in a large cohort. *Front. Biosci. (Elite Ed)* **2**, 547–558 (2010).
6 Monaco, A. P., Bertelson, C. J., Liechti-Gallati, S., Moser, H. & Kunkel, L. M. An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. *Genomics* **2**, 90–95 (1988).
7 Ousterout, D. G., Kabadi, A. M., Thakore, P. I., Majoros, W. H., Reddy, T. E. & Gersbach, C. A. Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. *Nat. Commun.* **6**, 6244 (2015).
8 Blonden, L. A., Grootscholten, P. M., den Dunnen, J. T., Bakker, E., Abbs, S., Bobrow, M. *et al.* 242 breakpoints in the 200-kb deletion-prone P20 region of the DMD gene are widely spread. *Genomics* **10**, 631–639 (1991).
9 Love, D. R., England, S. B., Speer, A., Marsden, R. F., Bloomfield, J. F., Roche, A. L. *et al.* Sequences of junction fragments in the deletion-prone region of the dystrophin gene. *Genomics* **10**, 57–67 (1991).
10 McNaughton, J. C., Cockburn, D. J., Hughes, G., Jones, W. A., Laing, N. G., Ray, P. N. *et al.* Is gene deletion in eukaryotes sequence-dependent? A study of nine deletion junctions and nineteen other deletion breakpoints in intron 7 of the human dystrophin gene. *Gene* **222**, 41–51 (1998).
11 Nobile, C., Toffolatti, L., Rizzi, F., Simionati, B., Nigro, V., Cardazzo, B. *et al.* Analysis of 22 deletion breakpoints in dystrophin intron 49. *Hum. Genet.* **110**, 418–421 (2002).
12 Toffolatti, L., Cardazzo, B., Nobile, C., Danieli, G. A., Gualandi, F., Muntoni, F. *et al.* Investigating the mechanism of chromosomal deletion: characterization of 39 deletion breakpoints in introns 47 and 48 of the human dystrophin gene. *Genomics* **80**, 523–530 (2002).
13 Sironi, M., Pozzoli, U., Cagliani, R., Giorda, R., Comi, G. P., Bardoni, A. *et al.* Relevance of sequence and structure elements for deletion events in the dystrophin gene major hot-spot. *Hum. Genet.* **112**, 272–288 (2003).
14 Gualandi, F., Rimessi, P., Trabanelli, C., Spitali, P., Neri, M., Patarnello, T. *et al.* Intronic breakpoint definition and transcription analysis in DMD/BMD patients with deletion/duplication at the 5′ mutation hot spot of the dystrophin gene. *Gene* **370**, 26–33 (2006).
15 Miyazaki, D., Yoshida, K., Fukushima, K., Nakamura, A., Suzuki, K., Sato, T. *et al.* Characterization of deletion breakpoints in patients with dystrophinopathy carrying a deletion of exons 45–55 of the Duchenne muscular dystrophy (DMD) gene. *J. Hum. Genet.* **54**, 127–130 (2009).
16 Ankala, A., Kohn, J. N., Hegde, A., Meka, A., Ephrem, C. L. H., Askree, S. H. *et al.* Aberrant firing of replication origins potentially explains intragenic nonrecurrent rearrangements within genes, including the human DMD gene. *Genome Res.* **22**, 25–34 (2012).
17 Ishmukhametova, A., Khau Van Kien, P., Méchin, D. B., 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.* **20**, 1096–1100 (2012).
18 Chen, C., Ma, H., Zhang, F., Chen, L., Xing, X., Wang, S. *et al.* Screening of Duchenne muscular dystrophy (DMD) mutations and investigating its mutational mechanism in Chinese patients. *PLoS ONE* **9**, e108038 (2014).
19 Oshima, J., Magner, D. B., Lee, J. A., Breman, A. M., Schmitt, E. S., White, L. D. *et al.* Regional genomic instability predisposes to complex dystrophin gene rearrangements. *Hum. Genet.* **126**, 411–423 (2009).
20 Kidd, J. M., Graves, T., Newman, T. L., Fulton, R., Hayden, H. S., Malig, M. *et al.* A human genome structural variation sequencing resource reveals insights into mutational mechanisms. *Cell* **143**, 837–847 (2010).
21 Carsana, A., Frisso, G., Tremolaterra, M. R., Lanzillo, R., Vitale, D., Santoro, L. *et al.* Analysis of dystrophin gene deletions indicates that the hinge III region of the protein correlates with disease severity. *Ann. Hum. Genet.* **69**, 253–259 (2005).
22 Frisso, G., Carsana, A., Tinto, N., Calcagno, G., Salvatore, F. & Sacchetti, L. Direct detection of exon deletions/duplications in female carriers of and male patients with Duchenne/Becker muscular dystrophy. *Clin. Chem.* **50**, 1435–1438 (2004).
23 Esposito, G., Ruggiero, R., Savarese, M., Savarese, G., Tremolaterra, M. R., Salvatore, F. *et al.* Prenatal molecular diagnosis of inherited neuromuscular diseases: Duchenne/Becker muscular dystrophy, myotonic dystrophy type 1 and spinal muscular atrophy. *Clin. Chem. Lab. Med.* **51**, 2239–2245 (2013).
24 Carsana, A., Frisso, G., Tremolaterra, M. R., Ricci, E., De Rasmo, D. & Salvatore, F. A larger spectrum of intragenic short tandem repeats improves linkage analysis and localization of intragenic recombination detection in the dystrophin gene: an analysis of 93 families from southern Italy. *J. Mol. Diagn.* **9**, 64–69 (2007).
25 Baskin, B., Stavropoulos, D. J., Rebeiro, P. A., Orr, J., Li, M., Steele, L. *et al.* Complex genomic rearrangements in the dystrophin gene due to replication-based mechanisms. *Mol. Genet. Genom. Med.* **2**, 539–547 (2014).
26 Krawczak, M. & Cooper, D. N. Gene deletions causing human genetic disease: mechanisms of mutagenesis and the role of the local DNA sequence environment. *Hum. Genet.* **86**, 425–441 (1991).

27 Rabbitts, T. H., Forster, A. & Milstein, C. P. Human immunoglobulin heavy chain genes: evolutionary comparisons of C mu, C delta and C gamma genes and associated switch sequences. *Nucleic Acids Res.* **9**, 4509–4524 (1981).
28 Abeysinghe, S. S., Chuzhanova, N., Krawczak, M., Ball, E. V. & Cooper, D. N. Translocation and gross deletion breakpoints in human inherited disease and cancer I: Nucleotide composition and recombination-associated motifs. *Hum. Mutat.* **22**, 229–244 (2003).
29 Singh, G. B., Kramer, J. A. & Krawetz, S. A. Mathematical model to predict regions of chromatin attachment to the nuclear matrix. *Nucleic Acids Res.* **25**, 1419–1425 (1997).
30 Mladenov, E. & Iliakis, G. Induction and repair of DNA double strand breaks: the increasing spectrum of non-homologous end joining pathways. *Mutat. Res.* **711**, 61–72 (2011).
31 Mitsui, J., Takahashi, Y., Goto, J., Tomiyama, H., Ishikawa, S., Yoshino, H. *et al.* Mechanisms of genomic instabilities underlying two common fragile-site-associated loci, PARK2 and DMD, in germ cell and cancer cell lines. *Am. J. Hum. Genet.* **87**, 75–89 (2010).
32 Drmanac, R., Petrovic, N., Glisin, V. & Crkvenjakov, R. A calculation of fragment lengths obtainable from human DNA with 78 restriction enzymes: an aid for cloning and mapping. *Nucleic Acids Res.* **14**, 4691–4692 (1986).
33 Li, H. L., Fujimoto, N., Sasakawa, N., Shirai, S., Ohkame, T., Sakuma, T. *et al.* Precise correction of the dystrophin gene in duchenne muscular dystrophy patient induced pluripotent stem cells by TALEN and CRISPR-Cas9. *Stem Cell Rep.* **4**, 143–154 (2015).