



HAL
open science

Pure intronic rearrangements leading to aberrant pseudoexon inclusion in dystrophinopathy: a new class of mutations?

Mouna Messaoud Khelifi, Aliya Ishmukhametova, Philippe Khau van Kien, Serge Perelman, Jean Poujet, Mireille Claustres, Sylvie Tuffery-Giraud

► To cite this version:

Mouna Messaoud Khelifi, Aliya Ishmukhametova, Philippe Khau van Kien, Serge Perelman, Jean Poujet, et al.. Pure intronic rearrangements leading to aberrant pseudoexon inclusion in dystrophinopathy: a new class of mutations?. *Human Mutation*, 2011, 32 (4), pp.467. 10.1002/humu.21471 . hal-00621292

HAL Id: hal-00621292

<https://hal.science/hal-00621292>

Submitted on 10 Sep 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**Pure intronic rearrangements leading to aberrant
pseudoexon
inclusion in dystrophinopathy: a new class of mutations?**

Journal:	<i>Human Mutation</i>
Manuscript ID:	humu-2010-0517.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	23-Dec-2010
Complete List of Authors:	Messaoud Khelifi, Mouna; Université Montpellier 1, UFR Médecine; Inserm U827 ISHMUKHAMETOVA, Aliya; Université Montpellier 1, UFR Médecine; Inserm U827 Khau van Kien, Philippe; CHU Montpellier, Laboratoire de Génétique Moléculaire Perelman, Serge; Lenval Foundation, Children's Hospital POUJET, JEAN; Hôpital la Timone, Service de Neurologie et maladies neuromusculaires Claustres, Mireille; Université Montpellier 1, UFR Médecine; Inserm U827; CHU Montpellier, Laboratoire de Génétique Moléculaire Tuffery-Giraud, Sylvie; Université Montpellier 1, UFR Médecine; Inserm U827
Key Words:	DMD gene, intronic rearrangement, double-deletion mutation, pseudoexon, array-CGH, RNA analysis

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

**Pure intronic rearrangements leading to aberrant pseudoexon
inclusion in dystrophinopathy: a new class of mutations?**

Mouna Messaoud Khelifi^{1,2*}, Aliya Ishmukhametova^{1,2*}, Philippe Khau Van Kien³, Delphine Thorel³, Déborah Méchin³, Serge Perelman⁴, Jean Pouget⁵, Mireille Claustres^{1,2,3}, Sylvie Tuffery-Giraud^{1,2&}.

* these two authors contributed equally to this work.

¹Université Montpellier 1, UFR médecine, Montpellier, F-34000 France,

²INSERM, U827, Montpellier, F-34000 France,

³CHU Montpellier, Hôpital Arnaud de Villeneuve, Laboratoire de Génétique Moléculaire, Montpellier, F-34000 France,

⁴LENVAL Foundation - Children's Hospital, 57 avenue de la Californie, 06200 Nice, France

⁵Service de Neurologie et Maladies Neuromusculaires, Centre de Référence national pour les Maladies Neuromusculaires et la SLA, Hôpital la Timone, 264 boulevard St-Pierre, 13385 Marseille cedex 05, France.

&Corresponding author:

Sylvie Tuffery-Giraud

Laboratoire de Génétique Moléculaire et INSERM U827

IURC, Institut Universitaire de Recherche Clinique

641 Avenue du Doyen Giraud

34093 MONTPELLIER Cedex 5, France

Tel : 33 4 67 41 53 83, Fax : 33 4 67 41 53 65, E-mail : sylvie.tuffery@inserm.fr

Abstract

We report on two unprecedented cases of pseudoexon activation in the *DMD* gene resulting from pure intronic double-deletion events that possibly involve microhomology-mediated mechanisms. Array comparative genomic hybridization analysis and direct genomic sequencing allowed us to elucidate the causes of the pathological pseudoexon inclusion detected in the RNA of the patients. In the first case (Duchenne phenotype), we showed that the inserted 387-bp pseudoexon was originated from an inverted ~57kb genomic region of intron 44 flanked by two deleted ~52kb and ~1kb segments. In the second case (Becker phenotype), we identified in intron 56 two small deletions of 592 bp (del 1) and 29 bp (del 2) directly flanking a 166-bp pseudoexon located in a very close proximity (134 bp) to exon 57. The key role of the del 1 in pseudoexon activation was established by using splicing reporter minigenes. However the analysis of mutant constructs failed to identify *cis* elements that regulate the inclusion of the pseudoexon and suggested that other splicing regulatory factors may be involved such as RNA structure. Our study introduces a new class of mutations in the *DMD* gene that emphasizes the potential role of underdetected intronic rearrangements in human diseases.

Keywords: *DMD* gene, intronic rearrangement, pseudoexon, RNA analysis, double-deletion mutation, array-CGH.

Introduction

The Duchenne Muscular Dystrophy gene (*DMD*; MIM*300377) is the largest gene detected to date. It spans approximately 2.2 megabases of the X chromosome and encodes several transcripts alternatively generated from 79 exons and 7 promoters. The transcript variant Dp427m expressed in muscle lineages is nearly 14kb long and is one of the longest [Muntoni et al., 2003]. Consequently, more than 99% of the gene sequence is composed of non-coding sequences. Mutations in the *DMD* gene cause the dystrophinopathies, a collective term for Duchenne Muscular Dystrophy (DMD) (MIM#310200), Becker Muscular Dystrophy (BMD) (MIM#300376), and the rare X-linked dilated cardiomyopathy (MIM#302045). DMD is a severe and rapid progressive neuromuscular disorder with the onset of symptoms generally occurring between 3 and 5 years and early loss of ambulation between the ages of 9 and 10 years, whereas BMD is a clinically less severe form of the disease in which affected individuals remain ambulatory beyond the age of 16 years and a few may lead a normal or near-normal life [Emery, 2002]. DMD is caused by mutations that disrupt the reading frame leading to a complete loss of functional dystrophin in muscle. In contrast, BMD is typically associated with in-frame mutations that allow production of either a reduced amount of normal dystrophin or an altered but partially functional dystrophin protein [Monaco et al., 1988].

The most common changes in the *DMD* gene consist of large genomic deletions or duplications of one or more exons, which account for mutations in 43%-85% and 7%-11% of all patients, respectively [Dent et al., 2005; Tuffery-Giraud et al., 2009; Flanigan et al., 2009]. Over the past years, the development of new diagnostic techniques such as the Multiplex Amplifiable Probe Hybridization (MAPH) [White et al., 2002], or the Multiplex Ligation-dependent Probe

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Amplification (MLPA) technique [Lalic et al., 2005] covering all 79 exons, has allowed to detect gene dosage imbalance for each of the 79 exons and thus to accurately define the extent of genomic rearrangements. However all these techniques focused on coding regions, leaving mutations located deep in the introns undetected. Recently, the development of high-density microarray-based comparative genomic hybridization (array-CGH) has provided a powerful tool to explore the entire genomic region of the *DMD* gene for unrecognized large copy number variations (CNVs) as defined by rearrangements of more than 1kb and even smaller insertions/deletions (Indels) as defined by a size <1kb [Hedge et al., 2008; Bovolenta et al., 2008].

The ~30% remaining mutations consist of small lesions, which are evenly distributed across the *DMD* gene. The implementation of a semiautomated direct sequencing methodology of all exons along with flanking intronic sequences, and promoters has enabled efficient detection of these small lesions [Flanigan et al., 2003]. Alternatively, RNA-based methods proved to be successful to detect point mutations in the *DMD* gene [Tuffery-Giraud et al., 2004; Deburgrave et al., 2007] and diagnostically valuable in clinical practice to determine the outcome of splice-site mutations and/or to identify alternative splicing patterns that may account for exceptions to the reading frame rule [Kesari et al., 2008]. Moreover, the analysis of mRNA obtained from muscle biopsies made possible the recognition of a novel class of disease-causing mutations in introns that cause missplicing by inducing inclusion of intronic sequences as exons (pseudoxon inclusion) [Tuffery-Giraud et al., 2003; Gurvich et al., 2008]. As reported in other genes [Buratti et al., 2006], the vast majority of this type of mutations was found to strengthen preexisting weak cryptic splice sites or to create new splice sites.

1
2 In this study, we report a novel class of mutation for pseudoexon activation in the *DMD*
3 gene. We show that this missplicing event can occur in the context of pure intronic
4 rearrangements as illustrated by the double deletions detected in two unrelated patients. One was
5 located deep in intron 44 and coupled with an inversion of a large genomic region while the
6 second one consists of two small deletions directly flanking a pseudoexon in intron 56. We
7 performed minigene assays to provide evidence of the pathogenic role of the identified deletions
8 in intron 56 upon exonization of the intronic sequence, and to investigate whether the local
9 context plays a role in splicing regulation of the pseudoexon in the wild type and mutant context.
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

Materials and methods

Patients

Genetic and laboratory testing was performed in the probands under conditions established by the French law and appropriate written informed consents were collected. Patient 1 was referred to us at five years old because of manifestation of DMD with very high serum creatine phosphokinases (CK) levels. He had been adopted but had a compatible familial history since his mother was reported to suffer from myalgia and to have high serum CK levels.

Deleted: "biological"

There was no other familial information. A muscle biopsy was performed, and immunofluorescence (IF) staining with dystrophin antibodies (Dys-1, -2 and -3) was negative.

The patient was lost-to-follow-up until the last evaluation at the age of 19 when a poor motor evolution of the disease was noted. The wheelchair use was reported at 8 years of age.

Echocardiography revealed the beginning of a dilated cardiomyopathy with a left-ventricular hypokinesia and reduced ejection fraction (EF) (<50%) and spirometry diagnosed a mild restrictive respiratory insufficiency with a 62% forced vital capacity (FVC).

Patient 2 was a 30-year-old man who was first examined at the age of 8 because of fatigability.

He complained of neither muscle pain nor cramps. He showed enlarged calves and he had an increased level of serum CK (7.500 IU/l, normal <200 IU/l). The family history was negative, and his development was normal. Due to consistently elevated serum CK levels, a muscle biopsy was performed at the age 16. It revealed discrete dystrophic features, and immunohistochemical dystrophin analysis showed decreased and irregular sarcolemmal labeling with Dys2 and Dys1 antibodies. Western-blotting showed reduced amount of a normal-sized protein (about 25% of the control level). The patient showed no signs of muscle weakness during childhood and later in

1
2 early adulthood, and was able to participate in intensive sport activities. Echocardiography was
3
4 normal until 21, then the Left ventricular Ejection Fraction (LVEF) decreased below 50% (last
5
6 LVEF of 45% at age 26) and the patient was treated with angiotensin-converting enzyme (ACE)
7
8 inhibitors.
9

10 11 12 **Mutation analysis**

13
14 Genomic DNA from the patients was screened for deletion and/or duplication using MLPA
15
16 (Salsa MLPA kit P034/P035 DMD/Becker MRC-Holland; Amsterdam, Netherlands). The
17
18 dystrophin (or *DMD*) transcripts were analyzed as previously described [Tuffery-Giraud et al.,
19
20 2004]. Briefly, total RNA was isolated from a frozen muscle biopsy and full-length cDNA was
21
22 amplified as ten separate and partially overlapping fragments using the Access Quick RT-PCR
23
24 System (Promega, Charbonnières-les-Bains, France). The amplified products were subsequently
25
26 analyzed by electrophoresis on a 1.5% agarose gel and by the protein truncation test (PTT). For
27
28 patient 1, because of the absence of amplification of the cDNA fragment encompassing exon 43
29
30 to exon 51, additional primers were designed to amplify the region in three overlapping
31
32 fragments. Fragments of normal size were obtained except for the region located between the
33
34 junction of the exons 43/44 and the exon 46 (forward, 5'-CCGACAAGGGCGATTTGACA-3'
35
36 and reverse, 5'-CTTGACTGCTCAAGCTTTTCTTTAG-3'). Abnormal cDNA fragments
37
38 were sequenced using the Big Dye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems,
39
40 Courtaboeuf, France). Following the detection of inserted intronic sequence in the transcripts,
41
42 primers were designed to amplify the related genomic region (intron 44, forward, 5'-
43
44 TGTATTGTCTGCTTTCATAC-3' and reverse 5'-GTGCCTGTATGTTAATTGTGA-3'; intron
45
46 56, forward, 5'-TGGCTAAGGGAAATGTTGCT-3' and reverse, 5'-
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2 CAGAAGTTCCTGCAGAGAAA-3') by using the PCR Master Mix (Promega, Charbonnières-
3 les-Bains, France). The PCR products were then directly sequenced. Nucleotide numbering for
4 mutation reflects cDNA numbering with +1 corresponding to the A of the ATG translation
5 initiation codon of GenBank NM_004006.2 (www.hgvs.org/mutnomen). Nucleotide numbering
6 for X chromosome position is given accordingly to the Genbank NC_000023.9 and the Human
7 Genome reference sequence of NCBI build 36/hg18 Mar.2006 (<http://genome.ucsc.edu/>).
8
9
10
11
12
13
14
15

16 **Array Comparative Genomic Hybridization (array-CGH)**

17 We used a custom-designed 12X135K NimbleGen microarray format (Roche NimbleGen,
18 Madison, WI). It includes 42 000 probes spanning the entire 2.2Mb *DMD* gene sequences on
19 chromosome X: 30,992,828–33,317,388 and numerous internal controls on autosomal, X and Y
20 chromosome loci (> 50 000 probes). Average probes length is 60 bases (range: 45-70 bases) with
21 isothermal melting temperature (T_m) of 42°C across the array. The average inner-spacing
22 between probes is 10 bases in the 79 exons with their 100 bp intronic borders, and 7 promoters.
23 Probes are interspersed with an average outer-spacing of 10 bp in the introns and the 50 kb
24 upstream and downstream genomic regions of the 5' and 3' UTR. The experiments were carried
25 out according to the manufacturer's recommended protocol (Roche NimbleGen). Briefly, 1 µg of
26 patient and reference DNA samples were labeled with green (Cy3) and red (Cy5) cyanines
27 fluorescent dyes, respectively. The microarray slides were hybridized for 72 hours at 42°C, then
28 washed, dried, and scanned using Innoscan 700A (INOPSYS, Toulouse, France). Array-CGH
29 data were extracted and analyzed using the NimbleScan version 2.5 software and SignalMap
30 version 1.9 software. For determining each breakpoint sequence, oligonucleotide primers pairs
31 were designed with the help of the Primer3Plus on-line tool (<http://www.bioinformatics.nl/cgi->
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2 bin/primer3plus/primer3plus.cgi) using both proximal and distal 0.7 kb flanking regions
3
4 determined by the CGH-array analyses (list of primers available in Supp. Table S1). PCR were
5
6 done in Patient 1 and Patient 2 using the Qiagen LongRange PCR kit (Qiagen, Courtaboeuf,
7
8 France) or the PCR Master Mix (Promega), respectively. Amplified junction fragments were
9
10 sequenced using the Big Dye terminator version 1.1 Cycle Sequencing Kit.
11

12 13 14 ***In silico* analysis of DMD sequences**

15
16 BLAST program (<http://blast.ncbi.nlm.nih.gov/>) was used to search for the origin of the
17
18 inserted sequence detected in the mature dystrophin transcripts. [Splice site score predictions for](#)
19
20 [the pseudoexons were performed using the Human Splicing Finder \(HSF\) web interface \(version](#)
21
22 [2.4; http://www.umd.be/HSF/](http://www.umd.be/HSF/)), [which includes position weight matrices to calculate consensus](#)
23
24 [values \(CV\) and an algorithm for the calculation of the MaxEnt scores \[Desmet et al., 2009\].](#) To
25
26 investigate the sequence characteristics in the vicinity of the breakpoints (± 100 bp), we searched
27
28 for extended homologies by means of the BLAST program and interspersed repeat-element
29
30 content with the BLAT and the RepeatMasker tools in the UCSC genome browser program
31
32 (<http://genome.ucsc.edu>).
33
34
35
36

37 **Minigene constructs, transfections and RT-PCR**

38
39 In Patient 2, we carried out functional assays to evaluate the splicing mechanism of the
40
41 pseudoexon (PE) identified in intron 56. Briefly, the pSPL3 exon trapping vector was used using
42
43 the procedure described previously [Le Guédard-Méreuze et al., 2010]. For the PE-WT (wild-
44
45 type) and PE-MT (mutant) constructs, fragments corresponding to the PE and flanking regions
46
47 were amplified from control and patient genomic DNA, respectively and inserted into the *XhoI*
48

1
2 and *NheI* restriction sites of the pSPL3 vector. The mutant constructs PE-D2 and PE-ISE were
3
4 generated by PCR-based mutagenesis (Quick Change Site Directed Mutagenesis Kit, Stratagene,
5
6 La Jolla, CA) from the PE-WT and PE-MT, respectively while the constructs PE-D1, PE-D1-1,
7
8 PE-D1-2, PE-D1-3, PE-D1-A, PE-D1-B, PE-AmpR and PE-D50 were created using the overlap
9
10 extension method [Lee et al., 2010] (list of primers [and sequences of deleted fragments](#) available
11
12 in Supp. Table [S2](#)). Three independent transfections assays of the minigenes in HeLa cells were
13
14 performed. RNA extraction and reverse transcription (RT)-PCR reactions were accomplished as
15
16 reported before [Le Guédard-Méreuze et al., 2010]. The products were resolved on 1.5% agarose
17
18 gel and splicing patterns confirmed by sequencing. The proportion of PE-inclusion transcripts
19
20 was measured using the Quantity one (v. 4.6.5) software (Bio-Rad, Marnes-La-Coquette,
21
22 France).
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Results

Dystrophin transcripts analysis was performed in two MLPA-negative patients to search for a small lesion in the *DMD* gene. In both cases, the inclusion of an intronic sequence in the mature transcripts was identified as the cause of the disease in the patients.

Pseudoexon characterization (Figure 1)

Patient 1

All RT-PCR products were identical to control samples except a cDNA fragment encompassing exons [43-46](#) for which a RT-PCR product of higher molecular weight was detected. Sequencing of the cDNA fragment disclosed the presence of a 387-bp long sequence inserted between exon 44 and exon 45 leading to premature insertion of a termination codon in the mature mRNA, [of which origin was undetermined at the time of analysis in 1996. This case was re-evaluated recently.](#) Alignment of the inserted sequence against genome sequences indicated that the sequence derived from the *DMD* intron 44 (c.6438_6439ins6439-106,288_6439-106,674) and was in inverse orientation ([the nucleotide sequence of the insertion is available in Supp. Table S1](#)). Sequencing of a genomic fragment encompassing the pseudoexon (PE) in the patient failed to detect any nucleotide change in the adjacent genomic regions. We thus decided to use high-density oligonucleotide array-CGH targeted to the entire *DMD* gene to be able to explore the whole 248 kb intron 44. Array-CGH analyses identified two non-contiguous deletions of ~52-kb and ~1 kb within intron 44 ([Supp. Figure S1](#)). We designed a series of primers to amplify the junction fragments ([Supp. Table S1](#)). Because amplifications with primers faced inward failed to give any products and the inserted sequence in the transcripts was inverted, we hypothesized that

1
2 the entire region between the two intronic deletions might be inverted. The use of two forward
3
4 and two reverse primers coupled together yielded amplification products and sequencing
5
6 identified the intronic breakpoints at chrX: 31,969,241 and chrX: 32,079,215 in intron 44, the
7
8 entire 57,133 bp region situated between the two deletions of 51,889 bp (del 1) and 951 bp (del
9
10 2) being inverted (Fig. 1A). This inversion put good splicing signals ([acceptor splice site: HSF](#)
11 [CV=79.5%, MaxEnt score=6.60; donor splice site: HSF CV= 94.3%, MaxEnt score= 11.01](#)) in a
12
13 favourable configuration around the 387-bp sequence so that it could be recognized as an exon
14
15 during pre-mRNA splicing. A 4-bp insertion (ACAT) was present at the 5' breakpoint and 2-bp
16
17 of microhomology was identified at the 3' breakpoint (Table 1). Moreover, bioinformatics
18
19 analysis showed the presence of interspersed repetitive elements at the two breakpoints (Table
20
21
22 1).

23 24 25 26 *Patient 2*

27
28 The dystrophin transcripts analysis revealed the presence of two products for the cDNA region
29
30 spanning exons 56 to 58, one corresponding to the control and one of higher molecular weight
31
32 (Fig. 2B). Sequencing of the upper band identified an out-of-frame insertion of 166 nucleotides
33
34 between exons 56 and 57 ([sequence available in Supp. Table S1](#)), derived from intron 56 and
35
36 located only a short distance (134 bp) upstream from exon 57 (r.8390_8391ins8391-300_8391-
37
38 135) (Fig. 1B). The 166-bp sequence displayed strong predicted splicing signals at either end
39
40 ([acceptor splice site: HSF=91.28%, MaxEnt score=8.19; donor splice site: HSF CV=88.47%,](#)
41 [MaxEnt score=9.72](#)), making the probability of a mutation reinforcing these sites unlikely to
42
43 explain the pseudoexon (PE) activation. The first attempts to amplify the genomic region
44
45 including the PE and 200 bp of flanking intronic sequences failed thus, we re-designed a forward
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2 primer located about 1kb upstream of the PE, and used a reverse primer located within exon 57,
3
4 that was known to be present. A smaller PCR product than expected was obtained for the patient,
5
6 whose sequencing revealed the presence of two distinct intronic deletions, one of 592 bp (del 1)
7
8 and a second one of 29 bp (del 2), on each side of the PE [c.8391-73_101del;8391-326_917del]
9
10 (Fig. 1B). No other mutation was detected at the 5' and 3' splice sites making it likely that the
11
12 identified complex genomic rearrangement could be responsible for the out-of-frame insertion of
13
14 the intronic sequence in the mature transcripts. Sequence inspection revealed six base pairs
15
16 (ATTAGT) and four base pairs (CTTT) microhomologies at the junctions of del 1 and del 2,
17
18 respectively (Fig. 1B), the latter corresponding to Topoisomerase I recognition sites
19
20 [(G/C)(A/T)T] [Been et al., 1984]. An array-CGH analysis confirmed the presence of the
21
22 rearrangement in intron 56 but as a single deletion (because of the array resolution) ([Supp.](#)
23
24 [Figure S1](#)). It did not reveal any other changes within the *DMD* gene except a previously
25
26 described frequent CNV in intron 2 [Bovolenta et al., 2008] ([Table 1](#)). [The mother of Patient 2](#)
27
28 [was found to harbour an identical to her son's genomic rearrangement indicating that the non-](#)
29
30 [contiguous two-part deletion probably occurred as a single concerted event.](#) We verified the
31
32 absence of the two deletions in intron 56 in a panel of more than 298 ethnically-matched control
33
34 chromosomes (95% confidence to detect a variant with an allele frequency of 1%) to rule out that
35
36 the patient-associated complex rearrangements [could](#) be explained by benign unreported
37
38 "CNVs/Indels".

41 **A sequence upstream of the pseudoexon in patient 2 is important for its regulation**

42
43 The sequence of the PE and its flanking regions was PCR-amplified from genomic DNA of the
44
45 patient 2 (PE-MT) and a control (PE-WT) and cloned in a pSPL3 exon trapping minigene (Fig.
46
47 2A). We analyzed the splicing pattern of the minigenes following transient transfection into
48
49

1
2 HeLa cells. The PE-WT transcripts showed full pseudoexon exclusion (Fig. 2C). In contrast, the
3
4 PE-MT minigene produced two bands corresponding to the PE inclusion and to a splicing event
5
6 between the vector exons, a profile similar to that observed in the patient (Fig. 2B). To determine
7
8 the role of each deletion individually on PE activation, we constructed minigenes carrying either
9
10 one (PE-D1) or the other (PE-D2) deletion (Fig. 2A). Interestingly, we found that only the PE-
11
12 D1 construct containing the upstream 592-bp deletion (del 1) was able to promote PE inclusion
13
14 whereas a normal splicing pattern was obtained with the PE-D2 construct containing the
15
16 downstream 29-bp deletion (del 2) (Fig. 2C). To confirm further the crucial role of the 592-bp
17
18 sequence in PE repression in the wild-type context and pseudoexon activation when deleted, we
19
20 inserted a heterologous 592-bp sequence from the bacterial gene AmpR [Kuga et al., 2000] in
21
22 place of the del 1 (Fig. 2A). Transcripts analysis of the PE-AmpR minigene after transfection in
23
24 HeLa cells revealed a complete PE inclusion (Fig. 2D), arguing for a role of the upstream 592-bp
25
26 sequence in PE repression in normal context. We wondered whether this repression role was
27
28 attributable to specific regulatory elements such as Intronic Splicing Silencers (ISS) whose
29
30 function would be to negatively regulate the PE splicing. To provide clues to this hypothesis, we
31
32 sequentially deleted the 592-bp sequence and generated a series of five different constructs (PE-
33
34 D1-1, PE-D1-2, PE-D1-3, PE-D1-A and PE-D1-B) harboring a combination of deletions to
35
36 narrow the region containing the putative regulatory element ([Supp. Figure S2](#)). Surprisingly,
37
38 none of the truncated versions of the minigene allowed exonization of the PE. We then
39
40 concluded that the serial deletions we made had not removed important silencer elements and
41
42 that the deletion of the whole 592-bp sequence (del 1) was required for PE activation.
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Formatted: Font: 12 pt

Role of Splicing Regulatory Elements in the pseudoexon activation

We next sought to explore whether the activation of the PE may result from the creation of a new Intronic Splicing Enhancer (ISE) motif at the breakpoint junction of the del 1 that would explain why its inclusion was dependent on the presence of the complete 592-bp deletion. An *in silico* analysis with the Human Splicing Finder (HSF) software predicted that the del 1 creates two overlapping potential binding sites for the SC35 (score: 82.85) and SRp40 (score: 93.41) SR proteins (Fig. 3A). To assess their role in PE activation, we abolished the two ISEs by introducing a single point mutation. The resulting construct (PE-ISE) was transfected in HeLa cells, but the transcripts analysis did not reveal any change in the splicing pattern compared to the PE-MT construct (Fig. 3B). We also investigated whether the del 1 could bring the PE closer to an activating element. Indeed, the HSF analysis indicated the presence of numerous ISEs and potential branch points upstream of the del 1. We thus deleted a sequence of 50 bp upstream of the del 1 (PE-D50) to eliminate all potential regulatory elements, but no major effect upon PE inclusion (46% versus 54 %) was observed in the minigene assay (Fig. 3C). Taken together, these results argued against a role of splicing regulatory elements in the exonization of the PE in the mutated context.

Discussion

Although the *DMD* gene was one of the earliest genes to be identified, finding all the mutations affecting the gene has been challenging owing to its large size and complexity. The report of two unparalleled cases of pure intronic double-deletion leading to missplicing events substantiates this observation.

Over the past years, the development and implementation of new DNA-based diagnostic methods has facilitated and improved the detection of the wide variety of different mutations (deletions, duplications, triplications, small lesions, insertions of repetitive sequences, genomic inversions, complex alleles...) that occur in the *DMD* gene. Despite these major technological advances, it is apparent that DNA-based strategies cannot pick-up all mutations and RNA analyses are still required, especially to recognize pseudoexon mutations. Moreover, the mutation remains unidentified in a few ascertained cases of dystrophinopathy suggesting the existence of as yet unknown mutational mechanisms [Flanigan et al., 2009].

Pseudoexons are intronic sequences that are approximately the same length as exons (50bp-200 bp) with apparently viable donor, acceptor and branch splice sites but which are not normally spliced in the mature mRNA transcript. There is evidence that inclusion of many of these sequences, which are usually very abundant in the introns of most genes, is actively inhibited due to the presence of intrinsic defects in their composition, the presence of silencer elements or the formation of inhibiting RNA secondary structures [Dhir and Buratti, 2010]. A distinct class of pseudoexon sequences derives from the exonization of *Alu* elements, the most abundant transposed elements in the human genome, which have accumulated mutations during the course of evolution and became recognized by the splicing machinery as exons. Transposed

1
2 [elements play a major role in shaping mammalian genomes and are involved in numerous](#)
3 [genetic diseases \[Sela et al., 2007; Vorechovsky, 2010\]](#). About thirty cases of pathological
4 pseudoexon inclusion in the *DMD* gene are reported in locus specific databases
5 (<http://www.umd.be/DMD/> and <http://www.dmd.nl>), mainly resulting from single nucleotide
6 substitutions that function by strengthening preexisting splice sites or by creating new ones as
7 observed in other genes [Gurvich et al., 2008; Bovolenta et al., 2008; Takeshima et al., 2010;
8 Dhir and Buratti, 2010]. A few additional cases have been reported that involve the
9 rearrangement of genomic regions. Among them, two reside in intron and do not extend to
10 adjacent exons. They consist in one case of a 11 kb deletion in intron 11 identified in a X-linked
11 dilated cardiomyopathy patient and leading to the exonization of a [novel fusion *Alu* exon](#) [Ferlini
12 et al., 1998], and in the other case, of a small 18-bp deletion within intron 37 inducing the
13 incorporation of a 77 bp PE between exons 37 and 38 of the *DMD* gene [Bovolenta et al., 2008].
14
15

16 To our knowledge, these cases together with the herein unprecedented cases of double-
17 deletion mutations are the only four experimentally proven examples of pure intronic
18 rearrangements in the *DMD* gene that lead to PE activation. However the mutations reported
19 here differ from the previously reported ones as they involve double-deletions. These [two](#)
20 double-deletions greatly differ in size and genomic configuration as one occurred only a short
21 distance upstream of an exon and involves close to each other small deletions that may escape
22 detection by array-CGH ([Patient 2, intron 56](#)) while the other one occurred deep in an intron, and
23 involves distant large size deletions coupled with the inversion of the intervening 58-kb genomic
24 sequence (Patient 1, intron 44). Four cases of genomic inversions flanked by deletion/duplication
25 in the *DMD* gene are described in the literature [Cagliani et al., 2004; Bovolenta et al., 2008;
26 [Madden et al., 2009; Oshima et al., 2009](#)], [and it is worth noting that all of these events are](#)
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2 present in the major deletion hot-spot around exons 43 to 53. All of them involve coding regions
3
4 leading to the skipping of the exons included in the aberrations in all cases, and the creation of
5
6 novel exons in two cases [Cagliani et al., 2004; Madden et al., 2009].
7

8
9 The analysis of the deletion junction sequences in Patient 1 and Patient 2 revealed the
10 presence of specific signatures frequently associated with complex genomic rearrangements. In
11 particular, we noticed the presence of microhomologies (2-6 bp) in three of the four deletion
12 junctions and one had a 4-bp inserted sequence (Patient 1) suggesting that microhomology-
13 mediated processes may have contributed to these rearrangements. Different categories of
14 mutational mechanism have been reported to give rise to genomic rearrangements [Chen et al.,
15 2010]. They include Non-Homologous End Joining (NHEJ), the most prominent DNA repair
16 mechanism, which is divided into two pathways, classical and non-classical, originally termed
17 Microhomology-Mediated End Joining (MMEJ). In NHEJ, the presence of terminal
18 microhomologies (typically 1-4 bp) facilitates classical NHEJ but is not absolutely necessary. By
19 contrast, the NHEJ junctions of two incompatible ends of the same Double-Strand Breaks (DSB)
20 are often characterized by small (typically 1-4 bp) deletions and/or insertions as seen in Patient 1.
21 Recently, replication-based mechanisms have been proposed to account for the multiple
22 breakpoints involved in complex rearrangements. Different models assuming serial replication
23 slippage have been described in particular the Fork Stalling and Template Switching (FoSTeS),
24 the Microhomology-Mediated Break-Induced Replication (MMBIR) [reviewed in Zhang et al.,
25 2009; Chen et al., 2010], or the more recently proposed Synthesis-Dependent MMEJ model (SD-
26 MMEJ) [Yu and McVey, 2010]. FoSTeS, MMBIR and SD-MMEJ are consistent with the
27 features of complexity (deletion/inversion) and microhomology at the junctions that are reported
28 here; although MMBIR is break-induced (*i.e.* generated by a collapsed replication fork), FoSTeS
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2 is initiated by replication fork stalling (*i.e.* no double-strand break is required). Our observations
3
4 are in line with a recent proposed model, which postulates that mitotic events, rather than meiotic
5
6 events, would play an important role in the formation of rare pathogenic CNVs [Vissers et al.,
7
8 2009]. Nevertheless, it remains a technical challenge to determine which exact mechanism is
9
10 responsible for selected disease-associated rearrangements. Moreover, other genomic
11
12 architectural features that may have contributed to the deletion formation were present at the
13
14 deletion breakpoints and consisted of repetitive elements (LINE, DNA repeats) in Patient 1 and
15
16 topoisomerase I cleavage sites in Patient 2, known to promote illegitimate recombination [Zhu
17
18 and Schiestl, 1996].
19

20 Whether the genomic inversion could reasonably be considered as the cause of the PE
21
22 activation in Patient 1, the molecular mechanisms underlying the recognition of the novel exon
23
24 in Patient 2 were unclear. Strikingly, this PE located only 134 bp upstream from exon 57, is
25
26 ignored by the splicing machinery in normal conditions despite having splice-site strength scores
27
28 higher than the calculated average scores for *DMD* exons (CV/MaxEnt scores of 91.28/8.19 vs
29
30 mean 3'ss scores of 86.26/8.13, and 88.47/9.72 vs mean 5'ss scores of 86.99/8.25). In an attempt
31
32 to clarify the involved mechanisms, we used splicing reporter minigenes to investigate whether
33
34 the two intronic deletions identified in Patient 2 have removed or created Splicing Regulatory
35
36 Elements (SREs). Our findings indicated that only the upstream 592-bp sequence (del 1) is
37
38 decisive for the PE inclusion and that the presence of this specific 592-bp sequence is required
39
40 for the PE repression in the wild-type context. Indeed, its replacement by a heterologous
41
42 sequence (AmpR) induced complete PE inclusion. Nevertheless, deletion mutants did not allow
43
44 to identify candidate silencer motifs within the 592-bp sequence. We could also ruled out the
45
46 hypothesis that the deletion has created a new enhancer splicing element at the deletion junction
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2 or brought a favorable branch point sequence within proximity of the pseudoexon 3' splice site,
3
4 provided that the potential branch-point sequences have been correctly predicted. The DMD
5
6 intron 56 PE is one of the few examples where pseudoexon inclusion occurs without changing
7
8 the splice sites directly. We could not identify here how the upstream deletion (del 1) exerts its
9
10 pathogenicity upon PE splicing. Based on our minigene experiments, SREs seem unlikely to play
11
12 a role in the activation event, even though we have to consider the possibility that the minigene
13
14 constructs used were not appropriate to evidence this putative splicing element. Despite
15
16 extensive efforts were made to elucidate the splicing code [Barash et al., 2010], the factors that
17
18 drive splicing decisions and allow differentiation of exons from long flanking introns are far
19
20 from being understood. Pre-mRNA secondary structure is increasingly recognized as a general
21
22 modifier of splicing events, and in particular would play a role in helping the splicing machinery
23
24 to distinguish between real exons and pseudoexons sequences [Buratti et al., 2007]. Conserved
25
26 stem-loop regions within introns can regulate donor-site usage and splicing efficiency as reported
27
28 for ATM and CFTR pseudoexons [Buratti et al., 2007] or for tau exon 10 alternative splicing
29
30 [Donahue et al., 2006]. Stem loop variants that destabilize this structure result in increased
31
32 splicing of tau exon 10 and contribute to neurodegenerative disorders [Liu and Gong, 2008]. In
33
34 the DMD gene, the skipping of the dystrophin Kobe exon 19 which has an 52 bp intra-exon
35
36 deletion near the 5' splice donor site has been attributed to the loss of a hairpin structure in the
37
38 truncated exon which prevent the splicing machinery from recognizing the splice sites [Matsuo
39
40 et al., 1992]. Evaluating the influence of RNA secondary structure on the processing of
41
42 individual exons is considered as a difficult task. In silico predictions may help in determining
43
44 the impact of mutations on RNA structure, but this approach is more challenging in presence of
45
46 complex rearrangements such as the deletions identified in Patient 2 than for single base
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2 substitutions. MFold predictions [Zucker, 2003] with flanking intronic sequences of different
3 sizes around the PE gave rise to a large number of different models for the *DMD* intron 56 PE.
4 We did not observe major differences in the accessibility of the PE donor and acceptor splice
5 sites between the wild-type and mutant context, but rather the analyses showed differences in the
6 structure of the pseudoexon itself (data not shown). However, no reliable working model of PE
7 recognition could be proposed. Besides RNA secondary structure, other factors would be
8 involved in splicing regulation. Recent data suggest that DNA structure in terms of nucleosome
9 positioning and specific histone modifications, which have a well established role in
10 transcription, may also have a role in splicing [Schwartz and Ast, 2010].
11
12
13
14
15
16
17
18
19

20 Our study reiterates the importance to combine DNA- and RNA-based approaches to
21 detect all kind of mutations in a gene, and in particular in the huge *DMD* gene whose mutational
22 spectrum is of unparalleled complexity. We demonstrate here that pure intronic rearrangements
23 could represent a new class of disease-causing mutations by inducing missplicing events. They
24 are non-detectable by the PCR-based methods commonly used for molecular diagnosis, and can
25 escape detection by array-CGH analysis depending on their size. Furthermore, complex alleles
26 are known to occur within the *DMD* gene even though at a low frequency. Our data raise the
27 possibility that some affected individuals may carry undetected cryptic intronic rearrangements
28 that have an impact on splicing. This hypothesis could explain some exceptions to the reading-
29 frame rule. Most importantly, the screening of such rearrangements (by RNA studies and/or
30 array-CGH) should influence the inclusion criteria in the design of exon-skipping clinical trials.
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Acknowledgements

The authors wish to thank Sylvie Chambert and Céline Saquet for their excellent technical contribution, and Dr Sabrina Sacconi and Dr Véronique Humbertclaude for providing clinical information. We are also grateful to Christiane Branlant for useful advice about RNA secondary structure. This work was supported by grants from the Association Française contre les Myopathies (AFM) to ST-G. AI is a recipient of a postdoctoral fellowship from AFM (grant n° 14178).

For Peer Review

References

- 1
2
3
4
5 Barash Y, Blencowe BJ, Frey BJ. 2010. Model-based detection of alternative splicing signals.
6
7 *Bioinformatics* 26:i325-333.
8
- 9 Been MD, Burgess RR, Champoux JJ. 1984. Nucleotide sequence preference at rat liver and
10
11 wheat germ type 1 DNA topoisomerase breakage sites in duplex SV40 DNA. *Nucleic*
12
13 *Acids Res* 12:3097-3114.
14
- 15 Bovolenta M, Neri M, Fini S, Fabris M, TrabANELLI C, Venturoli A, Martoni E, Bassi E, Spitali P,
16
17 Brioschi S, Falzarano MS, Rimessi P, Ciccone R, Ashton E, McCauley J, Yau S, Abbs S,
18
19 Muntoni F, Merlini L, Gualandi F, Ferlini A. 2008. A novel custom high density-
20
21 comparative genomic hybridization array detects common rearrangements as well as deep
22
23 intronic mutations in dystrophinopathies. *BMC Genomics* 9:572.
24
- 25 [Buckler AJ, Chang DD, Graw SL, Brook JD, Haber DA, Sharp PA, Housman DE. 1991. Exon](#)
26
27 [amplification: a strategy to isolate mammalian genes based on RNA splicing. *Proc Natl*](#)
28
29 [*Acad Sci U S A* 88:4005-4009.](#)
30
- 31 Buratti E, Baralle M, Baralle FE. 2006. Defective splicing, disease and therapy: searching for
32
33 master checkpoints in exon definition. *Nucleic Acids Res* 34:3494-3510.
34
- 35 Buratti E, Dhir A, Lewandowska MA, Baralle FE. 2007. RNA structure is a key regulatory
36
37 element in pathological ATM and CFTR pseudoexon inclusion events. *Nucleic Acids Res*
38
39 35:4369-4383.
40
- 41 Cagliani R, Sironi M, Ciafaloni E, Bardoni A, Fortunato F, Prella A, Serafini M, Bresolin N,
42
43 Comi GP. 2004. An intragenic deletion/inversion event in the DMD gene determines a
44
45 novel exon creation and results in a BMD phenotype. *Hum Genet* 115:13-18.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

[Chen JM, Cooper DN, Férec C, Kehrer-Sawatzki H, Patrinos GP. 2010. Genomic rearrangements in inherited disease and cancer. *Semin Cancer Biol* 20:222-233.](#)

Deburgrave N, Daoud F, Llense S, Barbot JC, Récan D, Peccate C, Burghes AH, Bérout C, Garcia L, Kaplan JC, Chelly J, Leturcq F. 2007. Protein- and mRNA-based phenotype-genotype correlations in DMD/BMD with point mutations and molecular basis for BMD with nonsense and frameshift mutations in the DMD gene. *Hum Mutat* 28:183-195.

Dent KM, Dunn DM, von Niederhausern AC, Aoyagi AT, Kerr L, Bromberg MB, Hart KJ, Tuohy T, White S, den Dunnen JT, Weiss RB, Flanigan KM. 2005. Improved molecular diagnosis of dystrophinopathies in an unselected clinical cohort. *Am J Med Genet A* 134:295-298.

Desmet FO, Hamroun D, Lalande M, Collod-Bérout G, Claustres M, Bérout C. 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 37:e67.

Dhir A, Buratti E. 2010. Alternative splicing: role of pseudoexons in human disease and potential therapeutic strategies. *FEBS J* 277:841-855.

[Donahue CP, Muratore C, Wu JY, Kosik KS, Wolfe MS. 2006. Stabilization of the tau exon 10 stem loop alters pre-mRNA splicing. *J Biol Chem* 281:23302-23306.](#)

Emery AE. 2002. The muscular dystrophies. *Lancet* 359:687-695.

Ferlini A, Galie N, Merlini L, Sewry C, Branzi A, Muntoni F. 1998. A novel Alu-like element rearranged in the dystrophin gene causes a splicing mutation in a family with X-linked dilated cardiomyopathy. *Am J Hum Genet* 63:436-446.

Flanigan KM, von Niederhausern A, Dunn DM, Alder J, Mendell JR, Weiss RB. 2003. Rapid direct analysis of the dystrophin gene. *Am J Hum Genet* 72: 931-939.

- 1
2 Flanigan KM, Dunn DM, von Niederhausen A, Soltanzadeh P, Gappmaier E, Howard MT,
3
4 Sampson JB, Mendell JR, Wall C, King WM, Pestronk A, Florence JM, Connolly AM,
5
6 Mathews KD, Stephan CM, Laubenthal KS, Wong BL, Morehart PJ, Meyer A, Finkel
7
8 RS, Bonnemann CG, Medne L, Day JW, Dalton JC, Margolis MK, Hinton VJ; United
9
10 Dystrophinopathy Project Consortium, Weiss RB. 2009. Mutational spectrum of DMD
11
12 mutations in dystrophinopathy patients: application of modern diagnostic techniques to a
13
14 large cohort. *Hum Mutat* 30:1657-1666.
- 15
16 Gurvich OL, Tuohy TM, Howard MT, Finkel RS, Medne L, Anderson CB, Weiss RB, Wilton
17
18 SD, Flanigan KM. 2008. DMD pseudoexon mutations: splicing efficiency, phenotype, and
19
20 potential therapy. *Ann Neurol* 63:81-89.
- 21
22 Hegde MR, Chin EL, Mulle JG, Okou DT, Warren ST, Zwick ME. 2008. Microarray-based
23
24 mutation detection in the dystrophin gene. *Hum Mutat* 29:1091-1099.
- 25
26 Kesari A, Pirra LN, Bremadesam L, McIntyre O, Gordon E, Dubrovsky AL, Viswanathan V,
27
28 Hoffman EP. 2008. Integrated DNA, cDNA, and protein studies in Becker muscular
29
30 dystrophy show high exception to the reading frame rule. *Hum Mutat* 29:728-737.
- 31
32 Kuga A, Okamoto R, Inoue M. 2000. ampR gene mutations that greatly increase class C beta-
33
34 lactamase activity in *Enterobacter cloacae*. *Antimicrob Agents Chemother* 44:561-567.
- 35
36 Lalic T, Vossen RH, Coffa J, Schouten JP, Guc-Scekic M, Radivojevic D, Djuriscic M, Breuning
37
38 MH, White SJ, den Dunnen JT. 2005. Deletion and duplication screening in the DMD
39
40 gene using MLPA. *Eur J Hum Genet* 13:1231-1234.
- 41
42 Lee J, Shin MK, Ryu DK, Kim S, Ryu WS. 2010. Insertion and deletion mutagenesis by overlap
43
44 extension PCR. *Methods Mol Biol* 634:137-146.
- 45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Le Guédard-Méreuze S, Vaché C, Baux D, Faugère V, Larrieu L, Abadie C, Janecke A, Claustres M, Roux AF, Tuffery-Giraud S. 2010. Ex vivo splicing assays of mutations at noncanonical positions of splice sites in USHER genes. *Hum Mutat* 31:347-355.

[Liu F, Gong CX. 2008. Tau exon 10 alternative splicing and tauopathies. *Mol Neurodegener* 3:8.](#)

[Madden HR, Fletcher S, Davis MR, Wilton SD. 2009. Characterization of a complex Duchenne muscular dystrophy-causing dystrophin gene inversion and restoration of the reading frame by induced exon skipping. *Hum Mutat* 30:22-28.](#)

[Matsuo M, Nishio H, Kitoh Y, Francke U, Nakamura H. 1992. Partial deletion of a dystrophin gene leads to exon skipping and to loss of an intra-exon hairpin structure from the predicted mRNA precursor. *Biochem Biophys Res Commun* 182:495-500.](#)

Monaco AP, Bertelson CJ, Liechti-Gallati S, Moser H, Kunkel LM. 1988. An explanation for the phenotypic differences between patients bearing partial deletions of the *DMD* locus. *Genomics* 2:90-95.

Muntoni F, Torelli S, Ferlini A. 2003. Dystrophin and mutations: one gene, several proteins, multiple phenotypes. *Lancet Neurol* 2:731-740.

Oshima J, Magner DB, Lee JA, Breman AM, Schmitt ES, White LD, Crowe CA, Merrill M, Jayakar P, Rajadhyaksha A, Eng CM, del Gaudio D. 2009. Regional genomic instability predisposes to complex dystrophin gene rearrangements. *Hum Genet* 126:411-423.

[Sela N, Mersch B, Gal-Mark N, Lev-Maor G, Hotz-Wagenblatt A, Ast G. 2007. Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome. *Genome Biol* 8\(6\):R127.](#)

[Schwartz S, Ast G. 2010. Chromatin density and splicing destiny: on the cross-talk between chromatin structure and splicing. *EMBO J* 29:1629-1636.](#)

Deleted: ¶

- 1
2 Takeshima Y, Yagi M, Okizuka Y, Awano H, Zhang Z, Yamauchi Y, Nishio H, Matsuo M.
3
4 2010. Mutation spectrum of the dystrophin gene in 442 Duchenne/Becker muscular
5
6 dystrophy cases from one Japanese referral center. *J Hum Genet* 55:379-388.
7
- 8 Tuffery-Giraud S, Saquet S, Chambert S, Claustres M. 2003. Pseudoexon activation in the DMD
9
10 gene as a novel mechanism for Becker muscular dystrophy. *Hum Mut* 21: 608-614.
11
- 12 Tuffery-Giraud S, Saquet C, Chambert S, Echenne B, Marie Cuisset J, Rivier F, Cossée M,
13
14 Philippe C, Monnier N, Bieth E, Recan D, Antoinette Voelckel M, Perelman S, Lambert
15
16 JC, Malcolm S, Claustres M. 2004. The role of muscle biopsy in analysis of the dystrophin
17
18 gene in Duchenne dystrophy: experience of a national referral centre.
19
20 *Neuromuscul Disord* 14:650-658.
21
- 22 Tuffery-Giraud S, Bérout C, Leturcq F, Yaou RB, Hamroun D, Michel-Calemard L, Moizard
23
24 MP, Bernard R, Cossée M, Boisseau P, Blayau M, Creveaux I, Guiochon-Mantel A, de
25
26 Martinville B, Philippe C, Monnier N, Bieth E, Khau Van Kien P, Desmet FO,
27
28 Humbertclaude V, Kaplan JC, Chelly J, Claustres M. 2009. Genotype-phenotype analysis
29
30 in 2,405 patients with a dystrophinopathy using the UMD-DMD database: a model of
31
32 nationwide knowledgebase. *Hum Mutat* 30:934-945.
33
- 34 Vissers LE, Bhatt SS, Janssen IM, Xia Z, Lalani SR, Pfundt R, Derwinska K, de Vries BB,
35
36 Gilissen C, Hoischen A, Nesteruk M, Wisniewiecka-Kowalnik B, Smyk M, Brunner HG,
37
38 Cheung SW, van Kessel AG, Veltman JA, Stankiewicz P. 2009. Rare pathogenic
39
40 microdeletions and tandem duplications are microhomology-mediated and stimulated by
41
42 local genomic architecture. *Hum Mol Genet* 18:3579-3593.
43
- 44 [Vorechovsky I. 2010. Transposable elements in disease-associated cryptic exons. *Hum Genet*](#)
45
46 [127:135–154.](#)
47

1
2 White S, Kalf M, Liu Q, Villerius M, Engelsma D, Kriek M, Vollebregt E, Bakker B, van
3
4 Ommen GJ, Breuning MH, den Dunnen JT. 2002. Comprehensive detection of genomic
5
6 duplications and deletions in the DMD gene, by use of multiple amplifiable probe
7
8 hybridization. *Am J Hum Genet* 71:365-374.
9

10 Yu AM, McVey M. 2010. Synthesis-dependent microhomology-mediated end joining accounts
11
12 for multiple types of repair junctions. *Nucleic Acids Res* 38:5706-5717.
13

14 Zhang F, Carvalho CM, Lupski JR. 2009. Complex human chromosomal and genomic
15
16 rearrangements. *Trends Genet* 25:298-307.
17

18 Zhu J, Schiestl RH. 1996. Topoisomerase I involvement in illegitimate recombination in
19
20 *Saccharomyces cerevisiae*. *Mol Cell Biol* 16:1805-1812.
21

22 [Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic*](#)
23
24 [Acids Res](#) 31:3406-3415.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Legends for Figures

Figure 1. Schematic representation of the rearrangements in the patients (not to scale).

(A) Organization of intron 44 in wild-type (WT) and patient 1 (Patient 1). *Patient 1, upper line:* scheme of the double-deletion (del 1 and del 2), indicated by scissors signs detected by the array-CGH. The distance between the different elements is given in base pairs (bp), exons are represented by gray boxes. The position/orientation of the primers is indicated in arrows: 1F,2F, forward primers and 1R,2R, reverse primers for the deletions 1 and 2, respectively; *Patient 1, lower line:* scheme of the rearrangement after breakpoint definition showing the inversion indicated by dashed arrows of the genomic region of intron 44 between the two deletions, and localization of the pseudoexon (black box). [The chromosomal position and sequence of the 387-bp PE are detailed in Supp. Table S1.](#) The insertion (ins) of the ACAT motif in the junction 1 is indicated by a triangle symbol. The AG and GT dinucleotides denote activated acceptor and donor splice sites, respectively ([the consensus value \(CV\) for the acceptor and donor splice sites is given in percentage as calculated by the HSF program. The corresponding MaxEnt scores are of 6.60 \(acceptor splice site\) and 11.01 \(donor splice site\).](#) [Repeated elements found out across the deletions are indicated.](#) (B) Organization of intron 56 in patient 2 (Patient 2). *Patient 2, upper line:* the scheme of the double-deletion (del 1 and del 2) in patient 2. The size of the deletions and the distance from the exons (gray boxes) to the PE (back box) are given in base pairs (bp). [The chromosomal position and sequence of the 166-bp PE are detailed in Supp. Table S1.](#) The AG and GT dinucleotides denote activated acceptor and donor splice sites, respectively ([the consensus value \(CV\) for the acceptor and donor splice sites is given in percentage as calculated](#)

Deleted: Repeated elements found out across the deletions also.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

by the HSF program. The corresponding MaxEnt scores are of 8.19 (acceptor splice site) and 9.72 (donor splice site). Patient 2, lower line: sequence context of the deletions (in lowercase in the boxes), and the sequence motifs around the junctions (in uppercase), showing microhomologies in bold characters. Topoisomerase I sites are indicated by lightning signs.

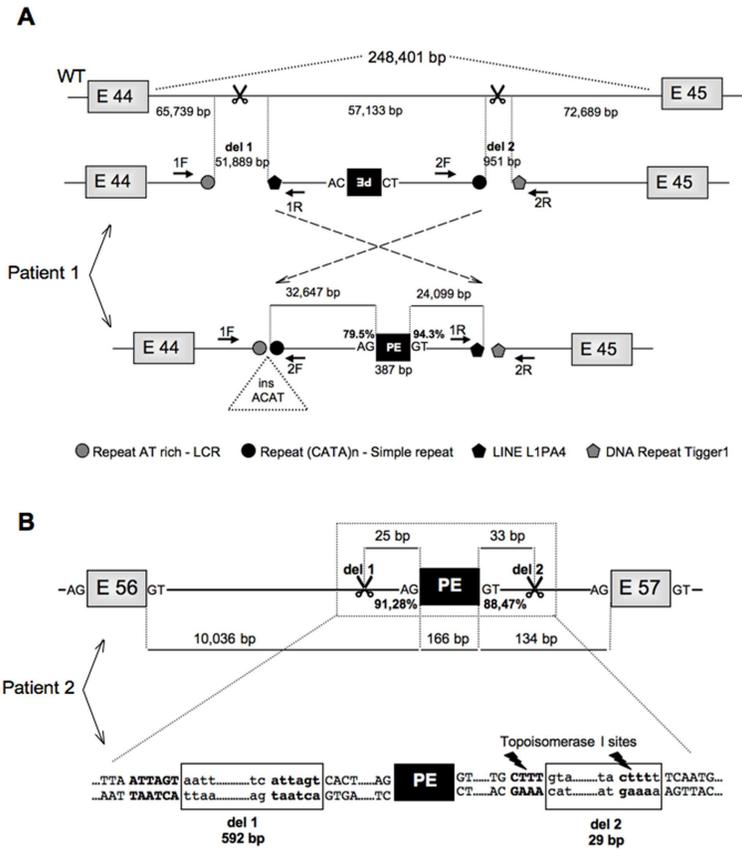
Figure 2. Role of the two adjacent intronic deletions in activation of the pseudoexon in DMD intron 56.

(A) Schematic representation of the heterologous three-exon, two-intron pSPL3 splicing reporter minigenes used in splicing assays and the subcloning of *DMD* intron 56 fragments isolated from wild-type (PE-WT) or mutant (PE-MT) alleles. The pSPL3 constructs contain an SV40 promoter, globin coding sequences (E1 and E2), HIV-1 *tat* splice donor (SD, MaxEnt score: 9.07) and acceptor (SA, MaxEnt score: 7.15) sequences compatible with splice sites from unrelated genes [Buckler et al., 1991], with the *DMD* pseudoexon (PE) as the middle exon (black boxes). (B) Reverse-transcription (RT)-PCR analysis of muscular dystrophin transcripts in Patient 2 (P2) showing the presence of a larger-sized product for the cDNA region spanning from exons 56 to 58 in addition to the normal-sized product obtained from the normal control (C). Sequencing of the normal sized product confirmed sequence normality. The identity of RNA products is shown on the right. (C, D) RT-PCR analysis using vector specific primers of transcripts derived from the indicated reporter minigenes following their expression in HeLa cells. (C) Note that only the minigene carrying the del 1 (PE-D1) induces pseudoexon insertion (PE insertion, PE+) at the same level as the PE-MT construct carrying the two deletions. The construct containing only the del 2 (PE-D2) gave rise to a normal splicing pattern (PE exclusion, PE-). (D) The replacement of the 592-bp sequence corresponding to the del 1 by a heterologous

1
2 sequence derived from the bacterial gene for ampicilline resistance (AmpR) was unable to
3 repress pseudoexon inclusion (100% PE inclusion, PE+). (*) An extra-band was detected with
4 this construct corresponding to the use of alternative splice sites located in the inserted AmpR
5 sequence. The identity of RNA products was established by sequencing of each band. Numbers
6 at the bottom of gels indicate the proportion (%) of PE inclusion (PE+) compared to normal
7 transcript (PE-). The percentages were determined using the Quantity one (v. 4.6.5) software.
8
9
10
11
12
13

14
15
16 **Figure 3. Role of splicing *cis*-acting elements in pseudoexon recognition.** (A) Schematic
17 representation of the pSPL3 minigene constructs used to investigate the role of *cis*-acting
18 splicing elements in the pseudoexon (black box) activation using the PE-MT construct carrying
19 the del 1 and del 2 deletions (scissors symbols). Two newly created splicing enhancer sequences
20 corresponding to SC35 and SRp40 binding motifs as predicted by the HSF program
21 (http://www.umd.be/HSF/, scores are given in brackets) were abrogated by site directed
22 mutagenesis (PE-ISE construct). The role of flanking *cis*-acting elements in PE activation, in
23 particular branch point sequences, was assessed by deleting a 50-nt sequence upstream the del 1
24 (PE-D50 construct). (B, C) PE-ISE and PE-D50 minigenes were used to transiently transfect
25 HeLa cells. After RNA isolation the splicing products were analyzed by RT-PCR using minigene
26 specific primers. No significant changes in the level of PE inclusion (PE+) was obtained
27 compared to the PE-MT construct (as defined in Fig. 2A). The PE exclusion rate (PE-) was
28 100% for the PE-WT construct. Numbers at the bottom of gels indicate the proportion (%) of PE
29 inclusion (PE+) compared to normal transcript (PE-). The percentages were determined using the
30 Quantity one (v. 4.6.5) software.
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



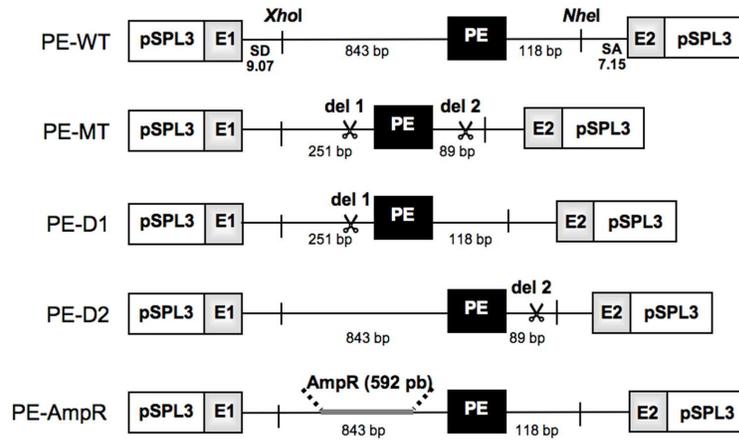
180x259mm (300 x 300 DPI)

Table 1. Characteristics of mutations detected in the patients.

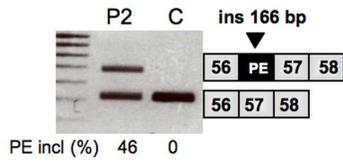
	Patient 1	Patient 2
Phenotype	DMD	Mild BMD
Multiplex PCR	Negative	Negative
MLPA	Negative	Negative
RT-PCR*	NM_004006.2: r.6438_6439ins6439-106,288_6439-106,674	NM_004006.2: r.8390_8391ins8391-300_8391-135
Array-CGH	Intron 44 deletion 1: NC_000023.9:g.(32,026,950-32,027,046)_(32,078,958-32,079,054)del Intron 44 deletion 2: NC_000023.9:g.(31,969,151-31,969,231)_(31,969,935-31,970,377)del	Intron 56: NC_000023.9:g.(31,425,001-31,425,017)_(31,425,801-31,425,911)del Intron 2: NC_000023.9:g.(32,897,108-32,897,190)_(32,898,216-32,898,708)del
Junction sequence	Junction 1: ins ACAT Junction 2: microhomology AA NC_000023.9:g.[31969242_31970192del951;32027326_32079214del51889insACAT;31970193_32027325inv57133]	Junction 1: microhomology ATTAGT Junction 2: microhomology CTTT NC_000023.9:g.[31425055_31425083del29;31425308_31425899del592]
Breakpoint findings	Junction 1: LCR: repeat AT-rich / Simple repeat: (CATA) _n , (AT) _n Junction 2: LINE: L1PA4 / DNA: Repeat Tigger 1	Unique sequence at the both sides of the deletions Topoisomerase I consensus cleavage site: CTT
Possible molecular mechanism	Non-homologous end joining (NHEJ) Microhomology-mediated replication-dependent recombination (MMRDR)	Non-homologous end joining (NHEJ) Microhomology-mediated replication-dependent recombination (MMRDR)

Abbreviations: PCR, Polymerase Chain Reaction; MLPA, Multiplex Ligation-dependent Probe Amplification; RT-PCR, ReverseTranscription-PCR; (*) the chromosomal positions and the nucleotide sequence of the pseudoexons are available in Suppl Table 1; array-CGH, array Comparative Genomic Hybridization; DMD, Duchenne muscular dystrophy phenotype; BMD, Becker muscular dystrophy phenotype; NM_004006.2 and NC_000023.9: accession numbers for DMD coding reference sequence and chromosome X reference sequence at NCBI Build 36.1 assembly (<http://www.ncbi.nlm.nih.gov>), respectively.

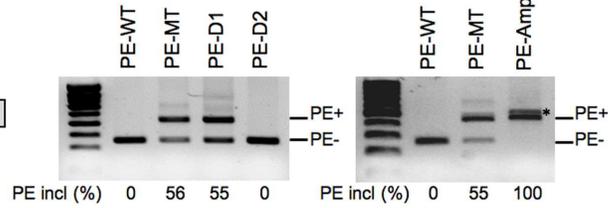
A



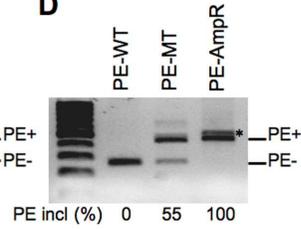
B



C

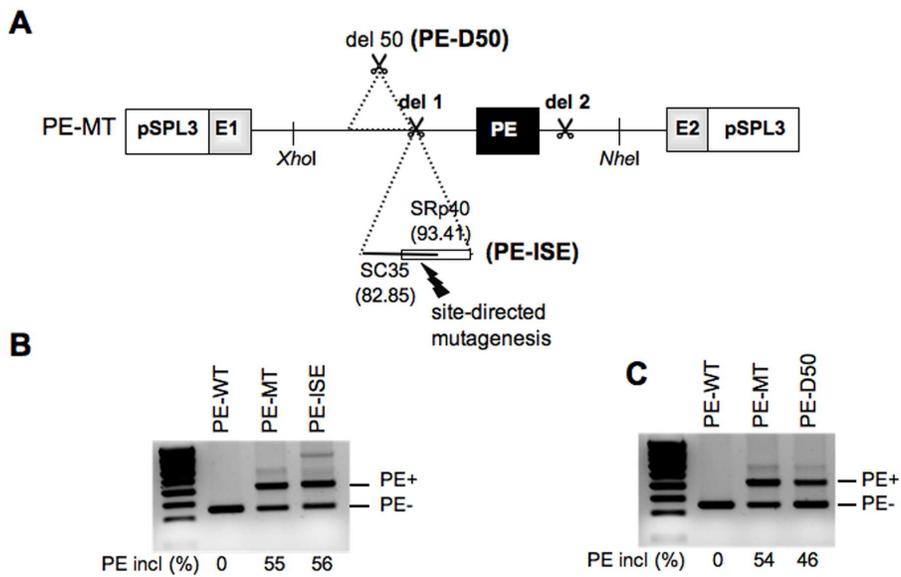


D



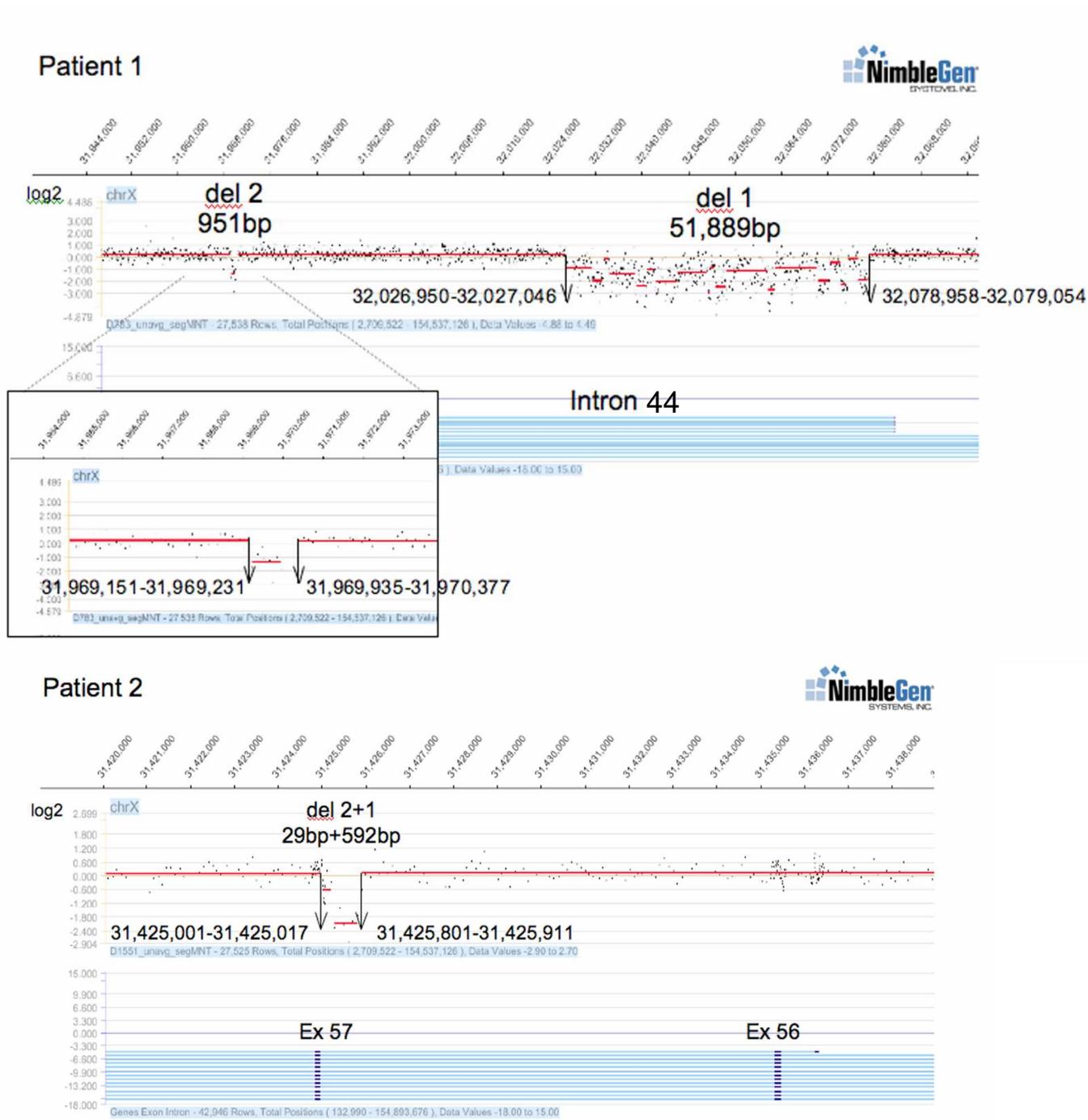
170x139mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



170x108mm (300 x 300 DPI)

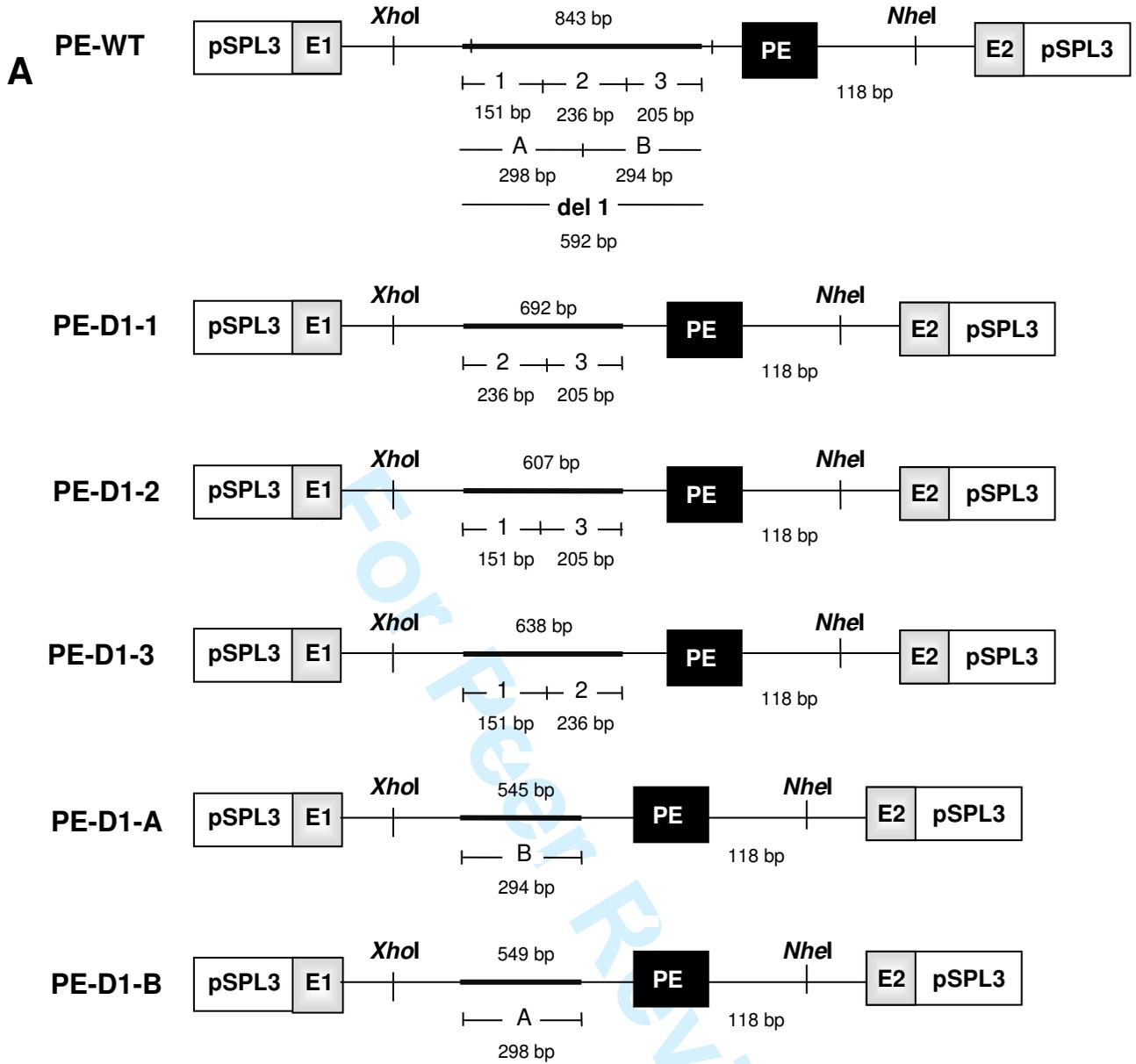
Supp. Figure S1



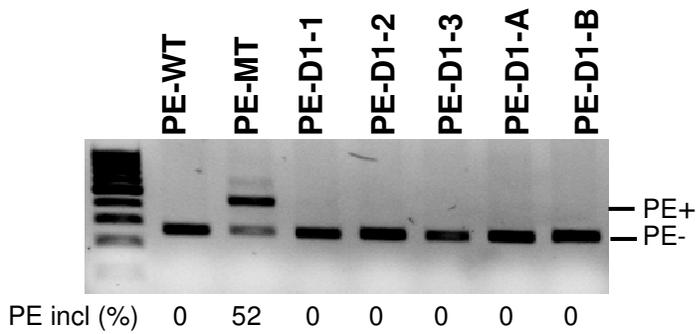
Legend for Supplementary Figure S1

The representative array results of DMD-CGH analysis using the segMNT algorithm available in the SignalMap version 1.9 software. The *DMD* gene coordinates on the X-chromosome are represented at the top (RefSeq NC_000023.9), with exon 1 to 79 from right to left. Scatter plots are shown for Patient 1 (top panel) carrying a double deletion in intron 44 and Patient 2 (low panel) carrying two distinct deletions of 29 bp and 592 bp in intron 56 visible as a single deletion by the array-CGH analysis.

Supp. Figure S2



B



Legend for Supplementary Figure S2**Search for silencer motifs within the 592-bp sequence upstream of the pseudoexon.**

(A) Schematic representation showing the strategy for 592-bp sequence serial deletion analysis. We generated a combination of five different deletions in the 592-bp sequence by overlap PCR using the pSPL3 minigene wild-type construct (PE-WT) to assess the role of *cis*-acting splicing elements that would repress pseudoexon (PE) inclusion in wild-type context and would be lost due to the 592-bp deletion (del 1) upstream of the PE (black box), thus activating the inclusion of the PE (PE+). The 592-bp region was deleted in three (PE-D1-1, PE-D1-2, PE-D1-3) or two (PE-D1-A, PE-D1-B) fragments and each of the minigenes was transiently transfected in HeLa cells. (B) After RNA isolation, the splicing products were analyzed by RT-PCR using minigene specific primers, and RT-PCR products were resolved on a 1.5% agarose gel. None of the deletion mutants allowed to activate PE inclusion (PE+) indicating that no crucial motif for PE repression has been deleted. Numbers at the bottom of gels indicate the proportion (%) of PE inclusion (PE+) compared to normal transcript (PE-). The percentages were determined using the Quantity one (v. 4.6.5) software.

Supp. Table S1 :

(A) Primers used to amplify and sequence the deletions junctions.

<i>Patients</i>	<i>Primers</i>	<i>Sequences</i>	<i>Chromosomal position</i>
Patient 1	1R	TGAAGTCAGGGCTCCACATT	X + 32026549_32026568
	2R	AAGAAGGGCAAAGGCAGATT	X + 31968704_31968723
	1F	TCCCTTGATGATTGTCACCTTG	X - 32079480_32079501
	2F	CCTTTCTACTATGCCCTTCACC	X - 31970376_31970397
Patient 2	F	CACGATTCAGTTCTTGGGAAA	X - 31426848_31426868
	R	GCCAAAAGAGATGGACGATT	X + 31424609_31424628

Abbreviations: X, X chromosome; "+" and "-", DNA strand +/-; F, forward primer; R, reverse primer.

(B) Chromosomal position and sequence of the pseudoexon identified in Patient 1 and Patient 2.

Patient 1 : 387-bp pseudoexon in intron 44. The 5' to 3' orientation of the *DMD* gene is on the complementary strand (Minus strand) on the X chromosome. Because the 387-bp insertion is on inverse orientation, it matched against the Plus strand of the reference sequence (NC_000023.9:g.32,002,840_32,003,226).

```
>[ref|NC_000023.9|NC_000023 Homo sapiens chromosome X, reference assembly, complete sequence
Length=154913754
```

```
Score = 715 bits (387), Expect = 0.0
Identities = 387/387 (100%), Gaps = 0/387 (0%)
Strand=Plus/Plus
```

```
Query 1      ATTCAGATGGACCCTAAGAGCTCAGTTACGGAAGTGGAAAGGGGAAAaacagcagcagca 60
Sbjct 32002840 ATTCAGATGGACCCTAAGAGCTCAGTTACGGAAGTGGAAAGGGGAAAACAGCAGCAGCA 32002899

Query 61      gcagcaacaacaacaacgacaacaacaacaacgacaacaacaacCGAAGAGGAAAAGAG 120
Sbjct 32002900 GCAGCAACAACAACAACGACAACAACAACGACAACAACAACCGAAGAGGAAAAGAG 32002959

Query 121     AGCTGAGTGGAAAAAGTCTTTGAAAATTTAGTCTTAAACTTATATTTACATTCTCATGC 180
Sbjct 32002960 AGCTGAGTGGAAAAAGTCTTTGAAAATTTAGTCTTAAACTTATATTTACATTCTCATGC 32003019

Query 181     TCTCCTGAAGTCACAGAAAGAGGTTTCGGAACCTTCTGCTTTAAGAGGACATTTGAATCGAC 240
Sbjct 32003020 TCTCCTGAAGTCACAGAAAGAGGTTTCGGAACCTTCTGCTTTAAGAGGACATTTGAATCGAC 32003079

Query 241     TGGAGAGGAACAAAGCATTGAAGAGAGAGGAAAGTGAACCTCCTCAGATAAAGCTCTAATAT 300
Sbjct 32003080 TGGAGAGGAACAAAGCATTGAAGAGAGAGGAAAGTGAACCTCCTCAGATAAAGCTCTAATAT 32003139

Query 301     GAGAAACATAGTTCCAGAGAGAAGTTAATGTTACTTCATTTTCTGTCTCATTTCCTTCA 360
Sbjct 32003140 GAGAAACATAGTTCCAGAGAGAAGTTAATGTTACTTCATTTTCTGTCTCATTTCCTTCA 32003199

Query 361     AGCTCAAAGCATGAGTAAGTGATAATG 387
Sbjct 32003200 AGCTCAAAGCATGAGTAAGTGATAATG 32003226
```

Patient 2 : 166-bp pseudoexon in intron 56. The 5' to 3' orientation of the *DMD* gene is on the complementary strand (Minus strand) on the X chromosome. The 166-bp insertion matched against the Minus strand of the reference sequence (NC_000023.9:g. 31,425,282_31,425,117).

> [ref|NC_000023.9|NC_000023](#) Homo sapiens chromosome X, reference assembly, complete sequence
Length=154913754

Score = 307 bits (166), Expect = 5e-83
Identities = 166/166 (100%), Gaps = 0/166 (0%)
Strand=Plus/Minus

```

Query 1      AAATATTAAGAATTGTTGACTACAACAGTATGGAAAAGCAATAGATTCCAGTGTGTATTT 60
             |||
Sbjct 31425282 AAATATTAAGAATTGTTGACTACAACAGTATGGAAAAGCAATAGATTCCAGTGTGTATTT 31425223
             |||
Query 61     CATGCCAAAAGTCTCAGCATTCTGCATGTGGAAATAAACATATGGCTAAACACTGCCTTT 120
             |||
Sbjct 31425222 CATGCCAAAAGTCTCAGCATTCTGCATGTGGAAATAAACATATGGCTAAACACTGCCTTT 31425163
             |||
Query 121    TCTCAAAATTGCCATCAAACACTATCCTCTGTTTTGTGGCTCTCAAAA 166
             |||
Sbjct 31425162 TCTCAAAATTGCCATCAAACACTATCCTCTGTTTTGTGGCTCTCAAAA 31425117
             |||

```

For Peer Review

Supp. Table S2: Primers used to generate the different reporter minigenes and corresponding deleted genomic sequences of the *DMD* intron 56.

<i>Minigenes</i>	<i>Primers</i>	<i>5' → 3' Sequence</i>
PE-WT	P-1-F	aattctggagctcgagcttcagaaagtttgaacaa
	P-2-R	ctcttaatttgctagcgcgtaccatgtcagaatc
PE-MT	P-1-F	aattctggagctcgagcttcagaaagtttgaacaa
	P-2-R	ctcttaatttgctagcgcgtaccatgtcagaatc
PE-D1	P-1-F	aattctggagctcgagcttcagaaagtttgaacaa
	P-3-R	gtagaaaaaattggatgcagtgactaattaattcataataacgc
	P-4-F P-2-R	cactgcatccaatttttctac ctcttaatttgctagcgcgtaccatgtcagaatc
PE-D2	P-5-F	agccagattttattcaagactgctttcaatggaattgttagaatc
	P-6-R	tcggctctaaaaataagtctgacgaaaagttaccttaacaatcttagtag
PE-AmpR	P-1-F	aattctggagctcgagcttcagaaagtttgaacaa
	P-7-R	cgacggggagtcaggcactaattaattcataataacgc
	P-8-F	gctgactccccgtcg
	P-9-R	gtagaaaaaattggatgcagtgcttttaagtctgctatgtg
	P-4-F	cactgcatccaatttttctac
	P-2-R	ctcttaatttgctagcgcgtaccatgtcagaatc
PE-D1-1	P-1-F	aattctggagctcgagcttcagaaagtttgaacaa
	P-10-R	aagaatacacaaatgaatgatctgtgactaattaattcataataacgc
	P-11-F P-2-R	cacagatcattcatttgtgtattctt ctcttaatttgctagcgcgtaccatgtcagaatc
PE-D1-2	P-1-F	aattctggagctcgagcttcagaaagtttgaacaa
	P-12-R	ctttccccctacccttctttaaagaatacacaaatgaatgatctgtg
	P-13-F P-2-R	taaagaagggtaggggaaag ctcttaatttgctagcgcgtaccatgtcagaatc
PE-D1-3	P-1-F	aattctggagctcgagcttcagaaagtttgaacaa
	P-14-R	gtagaaaaaattggatgcagtgctttccccctacccttcttta
	P-4-F P-2-R	cactgcatccaatttttctac ctcttaatttgctagcgcgtaccatgtcagaatc
PE-D1-A	P-1-F	aattctggagctcgagcttcagaaagtttgaacaa
	P-15-R	gtactaagacaacaactcacactaattaattcataataacgc
	P-16-F P-2-R	gtgagttgttcttagtac ctcttaatttgctagcgcgtaccatgtcagaatc
PE-D1-B	P-1-F	aattctggagctcgagcttcagaaagtttgaacaa
	P-17-R	gtagaaaaaattggatgcagtgtactaagacaacaactcac
	P-4-F P-2-R	cactgcatccaatttttctac ctcttaatttgctagcgcgtaccatgtcagaatc
PE-ISE	P-18-F	agcgttattatgaattaattagtea t ctgcatccaatttttctaccag
	P-19-R	tcgcaataataacttaattaatcagt a gacgtaggttaaaaaagatggtc
PE-D50	P-1-F	aattctggagctcgagcttcagaaagtttgaacaa
	P-20-R	gtagaaaaaattggatgcagtgttaacgctgcaagattgaaatg
	P-4-F P-2-R	cactgcatccaatttttctac ctcttaatttgctagcgcgtaccatgtcagaatc

PE-D1 : deletion of a 592-bp sequence

AATTGCTGATAATAGCTGAGTGATTGAGCATAATTTCTAATTTACCTGAAGATAAA
 GCTTTGCTAACACTGCGTTTCCTCTTTGTTTTCTGGATGATGATTTATTTTATTAATT
 TAGCTTCTCATCTTCAAGTCAAATGTGGATTTTATAGCACAGATCATTCAATTTGTGT
 ATTCTTAAATGGCTTCTAAGGATTAACGTGTTCTAAATACAGTTGACGGTAAAGCA
 CTCAGTCTCCTGCCTAAATTATTCATGTTTCGGGGGAAGCTTTCAGATAAATGTCTGA
 TTTTACTCTTCCATGTGAGTTGTTGTCTTAGTACTTTTTACACAAAGGAAACAAAGC
 AGAAAATGTTGAACTTGGTGAAGACAAATCCCAGGTGCACACAAATAAAGAAGG
 GTAGGGGGAAAGGAGACGCATTTGGGAAGAGGAGCAGAAAGGAACAGACGCCAG
 ATGGAAGAACTCAATGGAAAAGGCTGCCTAGGGTGTAGAAATGGAAAAGTCAAAA
 TGTGGGGAGAGACCTTTCCATTTCTCAAGGCAAAAAGAATTCCAGTACTAGCATGA
 GTCACATGAAAACGAAGTGTTTTTCATTAGT

PE-D2 : deletion of a 29-bp sequence

GTAGTTCACAATAGGTTTATTGTACTTTT

PE-AmpR : insertion of a 592-bp sequence

GCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCC
 CAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAA
 TAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGC
 CTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTA
 ATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTTCGT
 TTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCC
 CCCATGTTGTGCAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGTTGTCAGAAG
 TAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTAC
 TGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAAGTACTCAACCAAGTCATT
 CTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCCGGCGTCAATACGGGATA
 ATACCGCGCCACATAGCAGAACTTTAAAAG

PE-D1-1 :

AATTGCTGATAATAGCTGAGTGATTGAGCATAATTTCTAATTTACCTGAAGATAAA
 GCTTTGCTAACACTGCGTTTCCTCTTTGTTTTCTGGATGATGATTTATTTTATTAATT
 TAGCTTCTCATCTTCAAGTCAAATGTGGATTTTATAG

PE-D1-2 :

CACAGATCATTCAATTTGTGTATTCTTAAATGGCTTCTAAGGATTAACGTGTTCTAAA
 TACAGTTGACGGTAAAGCACTCAGTCTCCTGCCTAAATTATTCATGTTTCGGGGGAA
 GCTTTCAGATAAATGTCTGATTTTACTCTTCCATGTGAGTTGTTGTCTTAGTACTTTT
 TACACAAAGGAAACAAAGCAGAAAATGTTGAACTTGGTGAAGACAAATCCCAGG
 TGCACACAAA

PE-D1-3 :

TAAAGAAGGGTAGGGGGAAAGGAGACGCATTTGGGAAGAGGAGCAGAAAGGAAC
 AGACGCCAGATGGAAGAACTCAATGGAAAAGGCTGCCTAGGGTGTAGAAATGGAA
 AAGTCAAAATGTGGGGAGAGACCTTTCCATTTCTCAAGGCAAAAAGAATTCCAGTA
 CTAGCATGAGTCACATGAAAACGAAGTGTTTTTCATTAGT

PE-D1-A :

AATTGCTGATAATAGCTGAGTGATTGAGCATAATTTCTAATTTACCTGAAGATAAA
GCTTTGCTAACACTGCGTTTCCTCTTTGTTTTCTGGATGATGATTTATTTTATTAATT
TAGCTTCTCATCTTCAAGTGAAATGTGGATTTTATAGCACAGATCATTCATTTGTGT
ATTCTTAAATGGCTTCTAAGGATTAACGTGTTCTAAATACAGTTGACGGTAAAGCA
CTCAGTCTCCTGCCTAAATTATTCATGTTTCGGGGGAAGCTTTCAGATAAATGTCTGA
TTTTACTCTTCCAT

PE-D1-B :

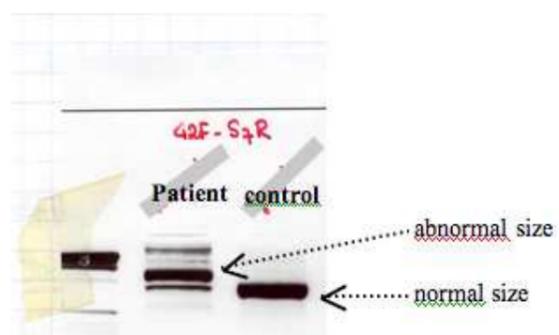
GTGAGTTGTTGTCTTAGTACTTTTTACACAAAGGAAACAAAGCAGAAAATGTTGAA
ACTTGGTGAAGACAAATCCCAGGTGCACACAAATAAAGAAGGGTAGGGGGAAAG
GAGACGCATTTGGGAAGAGGAGCAGAAAGGAACAGACGCCAGATGGAAGAAGTC
AATGGAAAAGGCTGCCTAGGGTGTAGAAATGGAAAAGTCAAAATGTGGGGAGAG
ACCTTTCCATTTCTCAAGGC AAAAAGAATTCCAGTACTAGCATGAGTCACATGAAA
ACGAAGTGTTTTTCATTAGT

PE-D50 :

AGTCTTTATTATATTTATGTTATAGGTAGCGTTATTATGAATTAATTAGT

For reviewers only, not for publication.

RT-PCR done in 1996 for Patient 1 showing the abnormal size band corresponding to the inclusion of the 387-bp PE between the exons 44 and 45.



For Peer Review