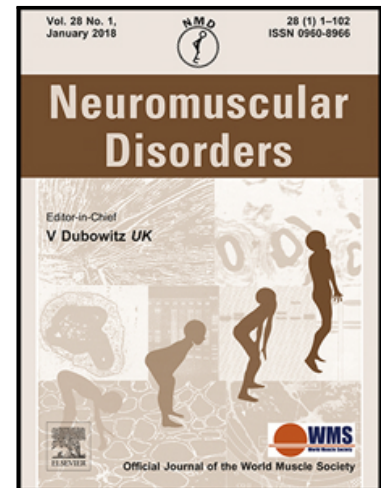# Journal Pre-proof

ANALYSIS OF COMPLEX STRUCTURAL VARIANTS IN THE DMD GENE IN ONE FAMILY

Leonela Luce , Martín M. Abelleyro , Micaela Carcione ,
Chiara Mazzanti , Liliana Rossetti , Pamela Radic , Irene Szijan ,
Sebastián Menazzi , Liliana Francipane , Julián Nevado ,
Pablo Lapunzina , Carlos De Brasi , Florencia Giliberto

Please cite this article as: Leonela Luce , Martín M. Abelleyro , Micaela Carcione , Chiara Mazzanti , Liliana Rossetti , Pamela Radic , Irene Szijan , Sebastián Menazzi , Liliana Francipane , Julián Nevado , Pablo Lapunzina , Carlos De Brasi , Florencia Giliberto , ANALYSIS OF COMPLEX STRUCTURAL VARIANTS IN THE DMD GENE IN ONE FAMILY, *Neuromuscular Disorders* (2020), doi: https://doi.org/10.1016/j.nmd.2020.11.015

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ANALYSIS OF COMPLEX STRUCTURAL VARIANTS IN THE *DMD* GENE IN ONE FAMILY

Leonela Luce[a,b,1], Martín M. Abelleyro[c,1], Micaela Carcione[a,b], Chiara Mazzanti[a,b], Liliana Rossetti[c], Pamela Radic[c], Irene Szijan[a], Sebastián Menazzi[d], Liliana Francipane[d], Julián Nevado[e,f,g], Pablo Lapunzina[e,f,g], Carlos De Brasi[c,2] and Florencia Giliberto[a,b,2]gilibertoflor@gmail.com

Affiliations:

[a]Universidad de Buenos Aires. Facultad de Farmacia y Bioquímica. Departamento de Microbiología, Inmunología, Biotecnología y Genética, Cátedra de Genética, Laboratorio de Distrofinopatías. Buenos Aires, Argentina.

[b]CONICET-Universidad de Buenos Aires. Instituto de Inmunología, Genética y Metabolismo (INIGEM). Buenos Aires, Argentina.

[c]CONICET-Academia Nacional de Medicina. Instituto de Medicina Experimental (IMEX). Buenos Aires, Argentina.

[d]Universidad de Buenos Aires. Hospital de Clínicas "José de San Martín". División de Genética. Buenos Aires, Argentina.

[e]Instituto de Genética Médica y Molecular (INGEMM)-IdiPAZ, Hospital Universitario La Paz, Universidad Autónoma, Madrid, Spain.

[f]Centro de Investigaciones Biomédicas en Red para Enfermedades Raras (CIBERER), Madrid, Spain.

[g]ITHACA- European Reference Network

[*]Corresponding author.

[1]LL and MMA have equally contributed to this paper and should be considered as first authors.

[2]FG and CDB have equally contributed to this paper and should be considered as last authors.

Highlights

- *DMD* complex structural variant molecular characterization by SNP-array and WGS

- Familial study showed mutational timeline, ancestral duplication and *de novo* deletion

- Bioinformatic analysis at breakpoints detects Double Strand Breaks stimulator motifs

- DNA "scars" as a clue to elucidate molecular mechanism underling structural variants

- Manifesting DMD woman showed skewed X-chromosome inactivation

Abstract

This work describes a family with Duchenne Muscular Dystrophy (DMD) with a rare case of a symptomatic pregnant woman. The main aim was to perform prenatal molecular diagnosis to provide genetic counseling. The secondary aim was to suggest the molecular mechanisms causing the complex structural variant (cxSV) identified. To accomplish this, we used a multi-technique algorithm including segregation analysis, Multiplex Ligation-dependent Probe Amplification, PCR, X-chromosome inactivation studies, microarrays, whole genome sequencing and bioinformatics. We identified a duplication of exons 38-43 in the *DMD* gene in all affected and obligate carrier members, proving that this was the DMD-causing mutation. We also observed a skewed X-chromosome inactivation in the symptomatic woman that explained her symptomatology. In addition, we identified a cxSV (duplication of exons 38-43 and deletion of exons 45-54) in the affected boy. The molecular characterization and bioinformatic analyses of the breakpoint junctions allowed us to identify Double Strand Breaks stimulator motifs and suggested the replication-dependent Fork Stalling and Template Switching as the most probable mechanisms leading to the duplication. In addition, the *de novo* deletion might have been the result of a germline inter-chromosome non-allelic recombination involving the Non-Homologous End Joining mechanism. In conclusion, the diagnostic strategy used allowed us to provide accurate molecular diagnosis and genetic counseling. In addition, the familial molecular diagnosis together with the in-depth characterization of the cxSV helped to determine the chronology of the molecular events, and propose and understand the molecular mechanisms involved in the generation of this complex rearrangement.

Pages: 26

1. Introduction

Dystrophinopathies are X-linked recessive diseases caused by pathogenic variants in the *DMD* gene (OMIM ID: 300377). Dystrophinopathies are characterized by mild to severe progressive degeneration and weakness of skeletal and cardiac muscles, and can be classified into: Duchenne Muscular Dystrophy (DMD), Becker Muscular Dystrophy (BMD) and DMD-associated Dilated Cardiomyopathy (DCM) [1,2]. DMD is the most prevalent pediatric form of muscular dystrophy, with an incidence of 1:3500-5000 male births [3]. The high rate of *de novo* mutations in the DMD gene (~33%) might explain the high incidence of families without previous history of the disease [4]. DMD is mainly caused by a complete absence of the dystrophin protein, which produces early muscle degeneration, leading to increased levels of serum Creatine Kinase (CK) [5].

Despite Dystrophinopathies are X-linked recessive diseases, some heterozygous female carriers develop symptoms, which may depend on X-chromosome inactivation patterns (XIP) [1]. Some reports have shown that most heterozygous females with skewed XIP develop signs of muscular dystrophy [6-9].

Identification of the disease-causing variant is crucial because it allows providing patients with the optimal standard of care and personalized mutation-dependent treatments. In addition, the molecular diagnosis provides information for genetic counseling. *DMD* molecular alterations comprise mainly Structural Variants (SVs) such as deletions (~68%) and duplications (~11%) of one or more exons, and small mutations (~20%) [10].

SVs include inversions, translocations, insertions and deletions of large DNA regions caused by errors during: Double Strand Breaks (DSB) repair pathways independent of DNA synthesis such as Non-Homologous End Joining (NHEJ) or Microhomology-Mediated End Joining (MMEJ); exchanges between highly homologous sequences during meiosis or DSB repair by Non-Allelic Homologous Recombination (NAHR); replication-dependent DSB repair pathways such as Break-Induced Replication (BIR) or Fork Stalling and Template Switching (FoSTeS); and retroposition of mobile elements [11–17].

The analysis of the sequences surrounding SV breakpoints, i.e. screening of DNA instability and DSB stimulators motifs, provide evidence about which molecular mechanism might have caused the SVs. Several heterogeneous motifs have been described in the literature: i) repetitive elements such as Alu, Long Terminal Repeats (LTRs) and Long Interspersed Nuclear Elements (LINEs), which stimulate DSB to initiate their transposition into elsewhere on the genome [18–21]; ii) Non-B DNA, such as tetraplex DNA, cruciform DNA, bent DNA, Z-DNA, and all sort of secondary structures [22,23]; iii) Short Tandem Repeats (STRs), which were found enriched at DSBs and in *DMD* intron breakpoint hotspots [24,25]; and iv) Alu/LINE specific retro-transposition target sequences [26–29].

Here, we present a familial case of DMD with a symptomatic pregnant woman and her affected nephew. Our main aim was to perform the prenatal molecular diagnosis of the unborn child. In addition, we undertook a thorough multi-technique molecular algorithm in order to identify the molecular mechanisms causing the complex SV (cxSV) found in this family.

2. Materials and methods

2.1. Patients

A familial DMD case was referred to our laboratory in 2005 (figure 1A). The index case was a 35-year-old symptomatic pregnant woman (II1) seeking prenatal diagnosis. She referred difficulties in walking and climbing stairs since she was 21 year-old. Laboratory results were: 3100 IU/l CK levels, normal female karyotype (46, XX) and a muscle biopsy compatible with DMD.

She had a family history of DMD with two relatives clinically diagnosed with the disease: a dead brother (II2) and a 17-year-old nephew (III4). The latter had undergone molecular diagnosis by end-point multiplex-PCR in 2005 that revealed a deletion of exons 45 to 54 in the *DMD* gene, resulting in a frame-shift mutation.

2.2. Samples

4

Whole blood was drawn by venipuncture in 5% ethylene-diamine tetraacetic acid (EDTA) as anticoagulant from the index case (II1), her sister (II3) and her affected nephew (III4) (Figure 1A). Genomic DNA (gDNA) was isolated using the cetyl-trimethyl-ammonium bromide (CTAB) method [30]. We also isolated gDNA from the fetus (III1, Figure 1A) Chorionic Villus Sample (CVS) using the QIAGEN DNeasy Blood and tissue kit [Redwood City, California (www.qiagen.com)]. DNA concentration and quality were measured by absorbance at 260nm and the ratios A260nm/A280nm and A260nm/A230nm, respectively.

The protocol was approved by the Institutional Review Board. Informed consent was obtained for all study subjects prior to the molecular studies.

2.3. Haplotype Assay

STRs analyses were designed in order to perform intrafamilial deletion segregation and to analyze CVS contamination with maternal blood. We selected and amplified STRs within (STR45 and STR49) and flanking (DYSII, STR44 and STR62) the deletion found in III4 [31–33]. The amplicons were labeled using 6-FAM-primers and PCR was performed as previously described, with minor modifications [34]. Primer sequences were obtained from the Leiden Muscular Dystrophy website [www.dmd.nl]. All PCR reactions were performed in a thermal cycler [Veriti; Applied Biosystems, Foster City, California]. The PCR products were analyzed using a fragment analyzer sequencer [ABI 3730XL; Applied Biosystems, Foster City, California] and data analysis was performed using the PeakScanner software [Applied Biosystems, Foster City, California].

2.4. Multiplex Ligation-dependent Probe Amplification (MLPA)

We used the MLPA kit for *DMD* (Salsas PO34-PO35) to detect deletions and duplications [35–37]. Reactions were performed according to the manufacturer's recommendations [MRC-Holland, Amsterdam, The Netherlands (www.mlpa.com)]. Products were analyzed using a fragment analyzer sequencer [ABI 3730XL; Applied Biosystems, Foster City, California] including the 500Liz size

standard. Data analysis was performed using Coffalyser [MRC-Holland, Amsterdam, The Netherlands] and GeneMarker V2.2.0 [Softgenetics, State College, Pennsylvania] softwares. Wild-type, deleted, and duplicated controls were included.

2.5. Human Androgen Receptor Assay (HUMARA)

This assay is based on a methylation-sensitive enzyme digestion followed by a PCR of an $(CAG)_n$ in the androgen receptor gene (*AR,* Xq12), and allows to study X-chromosome Inactivation Patterns (XIP). The protocol was performed as described by Allen *et al* [38]. *AR* allele profiles and areas under the curve were obtained from capillary electrophoresis analysis of PCR products using GeneMarker V2.2.0 software [Softgenetics, State College, Pennsylvania].

2.6. High density SNP-array

We conducted a genome-wide scan of 850,000 tag SNPs on III4 using Illumina CytoSNP-850k BeadChip. We quantified the DNA using PicoGreen [Invitrogen Corporation, Carlsbad, United States], and processed and hybridized the DNA to the chip according to manufacturer's specifications [Illumina, San Diego, United States]. Chip Image was processed and analyzed using the Chromosome Viewer tool in Genome Studio [Illumina, San Diego, United States]. Genotyping was estimated using the $\log_2R$ where R is the ratio of the observed normalized R-value for a SNP divided by the expected normalized R-value. GenCall scores <0.15 at any locus were considered as "no-calls". In addition, allele frequency was calculated for all SNPs. Genomic positions correspond to the Reference Genome GRCh37.

2.7. Whole genome sequencing (WGS) analysis

We sequenced the genome of III4 in a NovaSeq 6000 System [Illumina, San Diego, United States] by Macrogen services [Republic of Korea]. Bioinformatics analyses included a quality control with FASTX-toolkit (v.0.0.13.2), reads alignment to the Reference Genome (GRCh38.p12)

with Burrows-Wheeler Aligner tool (BWA v.0.7.15) and visualization of results with the Integrative Genomics Viewer software (IGV v.2.80).

2.8. Deletion Long Range-PCR (LR-PCR)

LR-PCR primers were designed flanking the 5' and 3' breakpoints of the deletion identified by the SNP-array (Supplementary table 1). Amplification products ranged between 2.0 Kb and 3.5 Kb. Reactions were performed following the recommendations of the enzyme's manufacturer [KAPA Biosystem, Wilmington, United States] in a BIOER thermal cycler [Hangzhou Bioer Technology, China]. PCR products were sequenced by Sanger reaction.

2.9. PCR amplification of duplication breakpoints

PCR amplification for the duplication breakpoints were based on the hypothesis of head-to-tail segmental fusion and adjusted by WGS results (primers described in Supplementary table 1). Reactions were performed following the recommendations of the enzyme's manufacturer [Promega, Madison, United States] in a thermal cycler [Biometra, Germany]. PCR products were sequenced by Sanger reaction.

2.10. Deletion and duplication breakpoints bioinformatic analysis

DNA intervals ranging between 10 bp and 50 bp centered on each 5' and 3' breakpoint of deletion and duplication were *in silico* screened for homologies, repetitive elements, non-B DNA, secondary structures and recombinogenic DNA motifs. These elements constitute a heterogeneous group of sequences that may act as stimulators for DSB, triggering an incorrect DNA repair/DNA replication leading to non-allelic recombination. The study was performed using the Human Reference Genome GRCh38 [NC_000023.11: 31641233-32372273 downloaded 5-Sep-2018 from the NCBI website, www.ncbi.nlm.nih.gov/] and was based on a recently reported strategy [39]. For this analysis, DSB stimulation motifs that showed significant Expected values (E-values <0.05) in

random points from the referred study were considered [39]. Bioinformatic analysis was mainly achieved using SeqBuilder and MegAlign programs [LaserGene DNA Star], ClustalW algorithm [www.ebi.ac.uk/Tools/msa/clustalw2/] and BLAST algorithm [blast.ncbi.nlm.nih.gov/Blast.cgi]. The RepeatMasker algorithm and Dfam [www.dfam.org/] were used to identify repetitive elements. Analysis of non-B DNA sequences was achieved by the non-B DNA motif search tool (nBMST) [nonb-abcc.ncifcrf.gov/apps/nBMST/default/] and confirmed by RepeatAround [portugene.com/repeataround.html] and QGRS mapper [bioinformatics.ramapo.edu/QGRS/analyze.php]. Secondary structure modelling was depicted using mfold [unafold.rna.albany.edu/?q=mfold]. Finally, among the recombinogenic motifs screened using SeqBuilder [LaserGene DNA Star], are included Scaffold Attachment Region (SAR), Ig heavy chain switch and hexanucleotide motifs targeted by the endonuclease/retro-transcriptase of mammalian retroposons (Jurka motifs) [29].

3. Results

3.1. Haplotype Assay

We performed STRs segregation analysis with the aims to confirm the absence of maternal blood contamination in the CVS and to identify the at-risk haplotype within this family. Haplotype analysis (figure 1A) allowed confirming the absence of maternal blood contamination since we detected only one maternal X-chromosome in the CVS. We were also able to identify two X-chromosomes in the CVS (one maternal and one paternal) suggesting a female fetus. Noteworthy, we also observed a 1-repeat retraction of STR49 in the fetus. In addition, as expected because of the previously detected deletion by multiplex-PCR, we observed amplification failure of STR45 and STR49 in III4 (figure 1A). We identified two alleles for STR49 in II1, II3 and III1; therefore, we concluded that they did not carry the deletion found in III4. Of note, they all share the 5'-end (STRDYSII: 222 and STR44: 191) of the at-risk haplotype (222/191/del/del/160). In addition, we

identified different paternal haplotypes for sisters II1 and II3. Finally, this analysis allowed us to detect a recombinant X-chromosome in III4 with a breakpoint between STR44 and STR62.

3.2. MLPA

Since, the STRs segregation results showed that the symptomatic female (II1) and the obligate carrier (II3) did not share the deletion of STR45 and STR49 with III4, we performed MLPA analysis in order to further characterize the molecular alteration in this family. Surprisingly, this analysis revealed a cxSV in III4. In addition to the deletion of exons 45-54 previously detected [NM_004006.3:c.(6382_6538)_(7909_8069)del], this analysis showed a duplication of exons 38-43 [NM_004006.3:c.(5290_5406)_(6196_6382)dup] (Figure 1B). The same duplication was also found in II1, II3 and III1 (data not shown). These results are in concordance with the haplotype assay described above.

3.3. HUMARA Assay

We studied XIP on II1 and II3. It is expected that, in average, ~50% of each X-chromosome is methylated (random inactivation) in females. Interestingly, we observed a skewed XIP of 96% in II1 (allele 21: digested/active, and allele 25: undigested/inactive), and random XIP of 65% in II3 (alleles 20/23). In addition, of note, sisters II1 and II3 did not share the maternal *AR* alleles (Figure 1C), suggesting another recombination event between *AR* and *DMD*. Also, we could corroborate the results of STRs analyses showing different paternal X-chromosomes.

3.4. Molecular characterization of the deletion by SNP-array, LR-PCR and WGS

The SNP-array analysis on III4 enabled us to estimate the 5' deletion breakpoint between rs1950112 (NC_00023.11:g.32095444) and rs1795577 (NC_00023.11:g.32092321), and the 3' breakpoint between rs2030002 (NC_00023.11:g.31646698) and rs5972426 (NC_00023.11:g.31646233) (Figure 2A). Therefore, the deletion spanned approximately 450 Kb.

9

In order to determine the exact points of rupture, we designed a LR-PCR to amplify and sequence the *loci* containing the breakpoints (Figure 2A). The primer pair AF1 and AR1 failed to amplify, suggesting that AR1 mapped within the deletion (Figure 2A). Then, we used a second reverse primer (AR2) that mapped 3' of AR1 which amplified a PCR product of approximately 2,435 bp (Figure 2A). The sequence of this amplicon allowed us to delimit the deletion that spanned 446,477 bp (NM_004006.3:c.6438+123812_8027+11362del).

Although the WGS was performed with the aim of characterizing the duplication (see below), it also allowed us to confirm the deletion breakpoints detected by LR-PCR, observed as several aligned chimeric reads (reads that did not align entirely on the reference sequence, such as ID6762, ID9251, ID26052) (Figure 2B, Supplementary Figure 1A).

Lastly, the analyses of the sequences showed a single-T microhomology between introns 44 and 54 (Figure 2C, Table 1).

3.5. Molecular characterization of the duplication by SNP-array and WGS

As shown above, the MLPA analysis revealed a duplication of exons 38-43 on III4. This duplication was also observed in the SNP-array. We were able to determine that the 5' breakpoint was between rs12856332 (NC_00023.11:g.32367273) and rs1801187 (NC_00023.11:g.32362879), and the 3' breakpoint between rs111931446 (NC_00023.11:g.32229089) and rs143786489 (NC_00023.11:g.32227327) (Figure 2A); spanning ~124 Kb. Because the distances between these pairs of SNPs were ~4 Kb and ~2 Kb, for the 5'and 3'breakpoints respectively, we were not able to design primers for PCR amplification and characterization of the breakpoints; not even under the head-to-tail fusion hypothesis (Figure 2A). Therefore, we performed a WGS with the aim to better identify the duplication breakpoints.

We observed chimeric reads comprised by sequences from introns 37 and 43, which mapped in the boundaries of a double depth-coverage region (e.g. ID22075, ID17613, ID18803, ID15702; Figure 2B). These results were consistent with a head-to-tail tandem duplication (Figure 2B,

10

Supplementary Figure 1B). Of note, we observed two discordances between SNP-array and WGS results. The SNP-array results showed a single copy of rs143786489 and double copy of rs1801187 (Figure 2A). However, the WGS results showed that rs143786489 was within and rs1801187 flanking the duplication (Figure 2B).

The sequence of the chimeric reads allowed us to design a duplication-specific head-to-tail PCR of 366 bp with the aim to confirm the duplication breakpoints by Sanger sequencing (Figure 2C). The size of the duplicated region was 131,284 bp.

Altogether, these data showed a 7 bp inverted insertion and a direct insertion of 11 bp, both insertions with intron 43 sequence identity (NM_004006.3:c.6291-5371_6291-5370ins[TAAAATGCAATTTCATTT;5326-5188_6291-5370]). This complex rearrangement suggests three events of template switching (TS) with microhomology of 1 bp "T" at first TS and a microhomology of 2 bp "TC" at third TS (Figure 2C, Table 1).

3.6. Bioinformatics analysis of SV breakpoints

The sequencing results described above allowed us to identify DSB stimulator motifs at all deletion and duplication breakpoints. The DNA motifs identified around each breakpoint are shown in Table 1. We found repetitive elements in 4/8 breakpoint junctions: one LTR and three LINE L1 (Table 1). In addition, 1/8 breakpoints presented a non-B DNA structure, a cruciform DNA motif found in the last 3' breakpoint of the duplication (Table 1).

We used the mfold algorithm to predict the folding of the DNA sequences surrounding each breakpoint junction. All sequences were predicted to fold into different secondary DNA structures with ΔG values ranging from -5.01 to -0.27, with an average of -1.77 Kcal (Figure 3). However, only 7/8 breakpoints were within the secondary DNA structure (Table 1, Figure 3).

Based on the hypothesis that the deletion resulted from a non-allelic non-homologous inter-chromosomal recombination assisted by events of ectopic synapsis between repeats, we further into the analysis of the breakpoint junctions with Dfam using a 5 Kb sequence. The Dfam algorithm

11

detected at least 8 partial- and full-length repetitive elements (LINEs, Alus, LTRs, MIRs, etc.) that might be involved in the formation of the ectopic synapsis (Figure 4). Furthermore, two in-tandem (TG)n and (T)n repeats were found, which might act as DNA breaks stimulators (Figure 4).

4. Discussion

Duchenne Muscular Dystrophy (DMD) is one of the most prevalent of the rare diseases [3]. It is caused by mutations in *DMD*, which is the largest gene of the human genome and maps in a genomic region with high recombination rates. These characteristics make *DMD* highly susceptible to mutation [40]. The current work presents the study of a family with DMD that included a rare case of a symptomatic female. We studied this family with the primary aim of providing molecular diagnosis and genetic counseling. Since we found a complex structural variant (cxSV) in one affected boy, we pursued the secondary aim of fully characterizing the cxSV and proposing the molecular mechanisms leading to this alteration. To accomplish these aims, we used a multi-technique and bioinformatic approach.

The index case was a DMD symptomatic pregnant woman seeking prenatal diagnosis (II1) with a nephew (III4, alive) and a brother (II2, dead) diagnosed with DMD. In 2005, we first did the molecular diagnosis of the DMD boy (III4) and we detected a deletion of exons 45-54 in *DMD* by end-point multiplex-PCRs. The deletion was later confirmed by STR and MLPA analyses.

For the prenatal diagnosis, in order to determine whether the CVS was contaminated with maternal cells, we performed a haplotype segregation study using STRs. This would allow us to analyze CVS contamination and, at the same time, corroborate the deletion in all studied family members. In this family, the STRs analysis allowed us to: 1) corroborate the deletion in III4; 2) detect heterozygosity of STR45 and STR49 in II1 and II3; 3) identify a recombination event in the *DMD* gene between STR44 and STR62 in III4; 4) identify a retraction of STR49 in III1; 5) identify different paternal X-chromosomes in II1 and II3; 6) confirm no-contamination of CVS with maternal cells; and 7) determine the genetic female gender of the fetus. The results presented here

remark that the analysis of intrafamilial STRs segregation, although a rather old molecular technique, is still a very useful tool.

Since this was a case of familial DMD, we had expected to find the deletion of exons 45-54 in II1 (symptomatic woman) and II3 (obligate carrier). However, the STR analyses failed to detect this deletion; therefore, we performed a MLPA analysis to further characterize the molecular alteration. Surprisingly, STR and MLPA analyses did not reveal the deletion in II1 and II3. However, the MLPA identified a second SV in the family, a duplication of exons 38-43. This duplication was found in all family members, proving that the duplication was the inherited disease-causing mutation. Therefore, the deletion initially found in III4 was re-categorized as a second mutational event. This second *de novo* alteration might have occurred during the early embryonic development of III4 or in his mother (II3) gametes (germline mosaicism). These results highlight the importance of retesting patients with *DMD* deletions originally diagnosed by end-point multiplex PCR using other complementary more sensitive techniques before including them in mutation-specific therapy protocols. Altogether, these methodologies might help to determine more precisely the deletion boundaries and/or to identify other SVs not detected by end-point multiplex PCR. The finding of other SVs would affect the patient eligibility and effectiveness of the mutation-specific therapies.

Since the family analyzed here presented with a rare case of a DMD-symptomatic woman, we wanted to look into the mechanism responsible for this phenotype. With the hypothesis that this woman had a skewed X-chromosome inactivation, we studied XIP by HUMARA. As expected, we observed a skewed XIP, suggesting that the active X-chromosome should be the chromosome carrying the duplication. In addition, also as expected, the asymptomatic woman (II3) showed a random XIP. Altogether, these results supported the hypothesis that the duplication and the skewed XIP were responsible for the DMD symptoms in II1. Moreover, as mentioned above, because the STRs analysis revealed that II1 and II3 had different paternal X-chromosomes we were able to identify the maternal X-chromosome as the one carrying the duplication. Of note, HUMARA

results showed that II1 and II3 did not share any *AR* allele. This confirmed the double paternity and suggested another recombination event between the mutated *DMD* (Xp21.2-p21.1) and *AR* (Xq12).

Although HUMARA has been recently validated on amniotic fluid for the prenatal risk prediction of DMD symptomatology in females [41], there are conflicting reports in the literature about the heritability of the XIP [9,42]. We proposed the parents to test for XIP in the fetus; however, as this study required a new amniocentesis procedure, they did not consent for this test. Therefore, we advised the parents to undergo XIP studies if III1 shows any clinical symptoms of DMD.

The molecular diagnosis of this family revealed a rare cxSV. In our laboratory, we have performed DMD molecular diagnosis in 437 patients and estimated a cxSV rate of 1.4% (6/437 DMD patients). The cxSVs found included: deletions-duplications, non-contiguous duplications and a large deletion with a 20 bp insertion. These results are similar to those reported by other authors where the occurrence of non-contiguous rearrangements within the same *DMD* allele had frequencies up to 2% [43].

Because molecular diagnosis usually involves the study of index cases without familial analysis, the molecular alterations identified are a picture of the final rearrangement at the time of diagnosis in the index case. The lack of familial genetic information is especially important when two or more mutational events (cxSVs) are detected. With this final picture, it is almost impossible to determine which of the mutations occurred first, and in which familial generation. The family studied herein is a clear example of how familial molecular diagnosis of two generations allowed us to determine a mutational timeline where the first mutational event was the duplication of exons 38-43 and the second event was the deletion of exons 45-54. The determination of the chronological order of these cxSVs might help to unravel molecular mechanisms where the first alteration becomes a predisposing factor for the second. This led us to the secondary aim of the study of proposing a molecular mechanism resulting in the observed cxSV.

First, we used a multi-technique approach (SNP-array, WGS and custom PCRs) to determine the breakpoints of the duplication and deletion. Once the breakpoints were identified, we performed a bioinformatic analysis of the sequences surrounding the SVs to seek DNA DSB stimulator motifs. These motifs have been found at breakpoint junctions more frequently than expected by chance according to Abelleyro *et al* [39].

The SNP-array allowed us to clearly identify the *loci* where the deletion breakpoints were located. This was particularly straightforward because patient III4 has one X-chromosome, and we were readily able to discriminate between no-signal/signal (deletion/single copy of *DMD*). The characterization of the deletion breakpoints was further confirmed with LR-PCR and Sanger sequencing. However, the identification of the duplication breakpoints was challenging because the SNP-array did not clearly differentiate between one and two gene copies. Therefore, we undertook a WGS approach to fully characterize the duplication breakpoints. The presence of chimeric reads allowed us to identify the breakpoints junctions (NM_004006.3:c.[6291-5371_6291-5370ins[TAAAATGCAATTTCATTT; 5326-5188_6291-5370];6438+123812_8027+11362del] - Leiden Open Variation Database (LOVD) v.3.0 URL: https://databases.lovd.nl/shared/individuals/00302913). Altogether, these results remark the importance of validating the SVs using several complementary methodologies.

After gathering all the detailed molecular information about the cxSV, we aimed to propose the molecular events that led to this cxSV. The tandem segmental duplication identified as the first mutational event, might have resulted from errors during MMBIR (Microhomology-Mediated Break-Induced Replication) or FoSTeS, both mechanisms involving *de novo* DNA synthesis. However, FoSTeS seemed to be more probable than MMBIR. This conclusion was based on the principle of maximum parsimony and the complexity of the event, characterized by at least three strand invasions preceded by replication fork collapses and facilitated by microhomologies (Supplementary Figure 2A). In addition, DNA strand break collapses that occur during these mechanisms might have been stimulated by the presence of Jurka hexa-nucleotides, secondary

structures and repetitive elements at relevant locations [44]. Furthermore, our recombination hypothesis relies on the presence of repetitive elements relatively near both deletion breakpoints, supporting a putative inter-chromosome non-allelic pairing structure or ectopic synapsis suitable for a localized recombination (Figure 4) [45]. Altogether, these proposed molecular mechanisms implied that II3 carries the *DMD* duplication in her germline DNA. This duplication resulted in a chimeric X-chromosome from both progenitors (I1 and I2) with an in-phase SV.

The second molecular event (deletion) that led to this cxSV might have occurred *de novo* during a meiotic recombination event. The maternal (II3) homologous X-chromosomes might have undergone recombination resulting in STRs alleles swap. The most likely scenario suggests that these two concomitant molecular events might had originated from a single inter-chromosomal recombination event that might had occurred either in II3's oocyte meiosis or in II3's oogonia mitosis, making the deletion inheritable. The detailed analysis of the deletion junction suggested a classical NHEJ because it did not have any molecular characteristic associated with a DNA replication dependent mechanism. We only identified a single-thymine of microhomology, which did not provide sufficient evidence to sub-classify the event as MMEJ (Supplementary figure 2B). In addition, also in support of the classical NHEJ model, we found DNA DSB stimulating motifs such as Jurka hexa-nucleotides and SAR at the 5' breakpoint and a Ig heavy chain switch region at the 3' breakpoint [29,46,47]. We also found other stimulators for DSB that might have been involved in the deletion (LTRs, STRs and highly stable secondary structures). In particular, long STRs might form non-B-DNA structures and make DNA susceptible to DSB.

Finally, we identified DNA secondary structures in 7/8 breakpoints. This finding highlights the important role of DNA secondary structures in the formation of DSB and SVs. This statement is supported by the fact that if we randomly select 50 bp of the genome it is very likely that they form secondary structures. However, if we simulate random breakpoints and analyze 25 bp sequences on each side, the chances of finding secondary structures are greatly reduced [39].

Overall, we were able to determine the chronology of the molecular events involved in the cxSV found in the studied family. We propose that the duplication might have aided to stabilize an unequal pairing between the duplicated X-chromosome and the non-duplicated homologue, and allowing for the approach of the regions involved in the deletion. This theory supports both previously mentioned hypotheses for the origin of the deletion. On one hand, the duplication could have helped the previously mentioned repetitive elements in the formation of the ectopic synapsis during female gametogenesis. While, it could have also been implicated in the unequal pairing that resulted in NHEJ mechanism.

In conclusion, the diagnostic strategy implemented in the present work allowed us to provide accurate molecular diagnosis and genetic counseling, aiding in the early diagnosis and selection of the appropriate standard-of-care and mutation-specific treatments. Finally, the exhaustive characterization of the SV breakpoints helped to propose and understand the molecular mechanisms involved in the generation of the cxSV.

mmc1.docx

mmc2.tif

mmc3.tif

References

[1] Darras BT, Urion DK, Ghosh PS. Dystrophinopathies. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, et al., editors. GeneReviews, Seattle (WA): University of Washington, Seattle; 2000.

[2] Brandsema JF, Darras BT. Dystrophinopathies. Semin Neurol 2015;35:369–84.

[3] Mendell JR, Shilling C, Leslie ND, Flanigan KM, al-Dahhak R, Gastier-Foster J, et al. Evidence-based path to newborn screening for Duchenne muscular dystrophy. Ann Neurol 2012;71:304–13.

[4] Haldane JBS. The rate of spontaneous mutation of a human gene. Journal of Genetics 2004;83:235–44. https://doi.org/10.1007/bf02717892.

[5] Emery AEH. Muscle histology and creatine kinase levels in the foetus in Duchenne muscular dystrophy. Nature 1977;266:472–3. https://doi.org/10.1038/266472a0.

[6] Pegoraro E, Schimke RN, Garcia C, Stern H, Cadaldini M, Angelini C, et al. Genetic and biochemical normalization in female carriers of Duchenne muscular dystrophy: evidence for failure of dystrophin production in dystrophin-competent myonuclei. Neurology 1995;45:677–90.

[7] Giliberto F, Radic CP, Luce L, Ferreiro V, de Brasi C, Szijan I. Symptomatic female carriers of Duchenne muscular dystrophy (DMD): genetic and clinical characterization. J Neurol Sci 2014;336:36–41.

[8] Viggiano E, Ergoli M, Picillo E, Politano L. Determining the role of skewed X-chromosome inactivation in developing muscle symptoms in carriers of Duchenne muscular dystrophy. Hum Genet 2016;135:685–98.

[9] Viggiano E, Picillo E, Ergoli M, Cirillo A, Del Gaudio S, Politano L. Skewed X-chromosome inactivation plays a crucial role in the onset of symptoms in carriers of Becker muscular dystrophy. The Journal of Gene Medicine 2017;19:e2952. https://doi.org/10.1002/jgm.2952.

[10] Aartsma-Rus A, Ginjaar IB, Bushby K. The importance of genetic diagnosis for Duchenne muscular dystrophy. Journal of Medical Genetics 2016;53:145–51. https://doi.org/10.1136/jmedgenet-2015-103387.

[11] Hastings PJ, Ira G, Lupski JR. A Microhomology-Mediated Break-Induced Replication Model for the Origin of Human Copy Number Variation. PLoS Genetics 2009;5:e1000327.

https://doi.org/10.1371/journal.pgen.1000327.

[12] Lee JA, Carvalho CMB, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell 2007;131:1235–47.

[13] Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW. The double-strand-break repair model for recombination. Cell 1983;33:25–35.

[14] Moore JK, Haber JE. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. Molecular and Cellular Biology 1996;16:2164–73. https://doi.org/10.1128/mcb.16.5.2164.

[15] McVey M, Lee SE. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends in Genetics 2008;24:529–38. https://doi.org/10.1016/j.tig.2008.08.007.

[16] McEachern MJ, Haber JE. Break-induced replication and recombinational telomere elongation in yeast. Annu Rev Biochem 2006;75:111–35.

[17] Quinlan AR, Hall IM. Characterizing complex structural variation in germline and somatic genomes. Trends Genet 2012;28:43–53.

[18] Rüdiger NS, Gregersen N, Kielland-Brandt MC. One short well conserved region ofAlu-sequences is involved in human gene rearrangements and has homology with prokaryoticchi. Nucleic Acids Research 1995;23:256–60. https://doi.org/10.1093/nar/23.2.256.

[19] Deininger PL, Batzer MA. Alu repeats and human disease. Mol Genet Metab 1999;67:183–93.

[20] Kazazian HH Jr. Mobile elements: drivers of genome evolution. Science 2004;303:1626–32.

[21] Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, et al. Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci U S A 2003;100:5280–5.

[22] Bacolla A, Wells RD. Non-B DNA Conformations, Genomic Rearrangements, and Human Disease. Journal of Biological Chemistry 2004;279:47411–4. https://doi.org/10.1074/jbc.r400028200.

[23] Wang G, Vasquez KM. Impact of alternative DNA structures on DNA damage, DNA repair, and genetic instability. DNA Repair 2014;19:143–51. https://doi.org/10.1016/j.dnarep.2014.03.017.

[24] Zavodna M, Bagshaw A, Brauning R, Gemmell NJ. The effects of transcription and recombination on mutational dynamics of short tandem repeats. Nucleic Acids Research 2018;46:1321–30. https://doi.org/10.1093/nar/gkx1253.

[25] Luce LN, Dalamon V, Ferrer M, Parma D, Szijan I, Giliberto F. MLPA analysis of an Argentine cohort of patients with dystrophinopathy: Association of intron breakpoints hot spots with STR abundance in DMD gene. J Neurol Sci 2016;365:22–30.

[26] Been MD, Burgess RR, Champoux JJ. Nucleotide sequence preference at rat liver and wheat germ type 1 DNA topoisomerase breakage sites in duplex SV40 DNA. Nucleic Acids Res 1984;12:3097–114.

[27] Spitzner JR, Muller MT. A consensus sequence for cleavage by vertebrate DNA topoisomerase II. Nucleic Acids Res 1988;16:5533–56.

[28] Weaver DT, DePamphilis ML. Specific sequences in native DNA that arrest synthesis by DNA polymerase alpha. J Biol Chem 1982;257:2075–86.

[29] Jurka J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc Natl Acad Sci U S A 1997;94:1872–7.

[30] Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Research 1980;8:4321–6. https://doi.org/10.1093/nar/8.19.4321.

[31] Clemens PR, Fenwick RG, Chamberlain JS, Gibbs RA, de Andrade M, Chakraborty R, et al. Carrier detection and prenatal diagnosis in Duchenne and Becker muscular dystrophy families, using dinucleotide repeat polymorphisms. Am J Hum Genet 1991;49:951–60.

[32] Beggs AH, Kunkel LM. A polymorphic CACA repeat in the 3′ untranslated region of dystrophin. Nucleic Acids Research 1990;18:1931–1931. https://doi.org/10.1093/nar/18.7.1931.

[33] King SC, Roche AL, Rita Passos-Bueno M, Takata R, Zatz M, Cockburn DJ, et al. Molecular characterization of further dystrophin gene microsatellites. Molecular and Cellular Probes 1995;9:361–70. https://doi.org/10.1016/s0890-8508(95)91700-4.

[34] Luce LN, Ottaviani D, Ferrer M, Szijan I, Cotignola J, Giliberto F. Molecular diagnosis of dystrophinopathies using a multi-technique analysis algorithm. Muscle & Nerve 2014;49:249–56. https://doi.org/10.1002/mus.23906.

[35] Schwartz M, Dunø M. Multiplex ligation-dependent probe amplification is superior for detecting deletions/duplications in Duchenne muscular dystrophy. Clinical Genetics 2004;67:189–91. https://doi.org/10.1111/j.1399-0004.2004.00382.x.

[36] Gatta V, Scarciolla O, Gaspari AR, Palka C, De Angelis MV, Di Muzio A, et al. Identification of deletions and duplications of the DMD gene in affected males and carrier females by multiple ligation probe amplification (MLPA). Human Genetics 2005;117:92–8. https://doi.org/10.1007/s00439-005-1270-7.

[37] Janssen B, Hartmann C, Scholz V, Jauch A, Zschocke J. MLPA analysis for the detection of deletions, duplications and complex rearrangements in the dystrophin gene: potential and pitfalls. Neurogenetics 2005;6:29–35. https://doi.org/10.1007/s10048-004-0204-1.

[38] Allen RC, Zoghbi HY, Moseley AB, Rosenblatt HM, Belmont JW. Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. Am J Hum Genet 1992;51:1229–39.

[39] Abelleyro MM, Radic CP, Marchione VD, Waisman K, Tetzlaff T, Neme D, et al. Molecular insights into the mechanism of nonrecurrent F8 structural variants: Full breakpoint characterization and bioinformatics of DNA elements implicated in the upmost severe phenotype in hemophilia A. Hum Mutat 2020;41:825–36.

[40] Nguyen Q, Yokota T. Antisense oligonucleotides for the treatment of cardiomyopathy in Duchenne muscular dystrophy. Am J Transl Res 2019;11:1202–18.

[41] He W-B, Du J, Xie P-Y, Zhou S, Zhang Y-X, Lu G-X, et al. X-chromosome inactivation

pattern of amniocytes predicts the risk of dystrophinopathy in fetal carriers of DMD mutations. Prenat Diagn 2019;39:603–8.

[42] Renault NK, Dyack S, Dobson MJ, Costa T, Lam WL, Greer WL. Heritable skewed X-chromosome inactivation leads to haemophilia A expression in heterozygous females. Eur J Hum Genet 2007;15:628–37.

[43] Kerr R, Robinson C, Essop FB, Krause A. Genetic testing for Duchenne/Becker muscular dystrophy in Johannesburg, South Africa. S Afr Med J 2013;103:999–1004.

[44] Vissers LELM, Lisenka E L, Bhatt SS, Janssen IM, Xia Z, Lalani SR, et al. Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genomic architecture. Human Molecular Genetics 2009;18:3579–93. https://doi.org/10.1093/hmg/ddp306.

[45] Liu P, Carvalho CMB, Hastings PJ, Lupski JR. Mechanisms for recurrent and complex human genomic rearrangements. Current Opinion in Genetics & Development 2012;22:211–20. https://doi.org/10.1016/j.gde.2012.02.012.

[46] Gale JM, Tobey RA, D'Anna JA. Localization and DNA sequence of a replication origin in the rhodopsin gene locus of Chinese hamster cells. Journal of Molecular Biology 1992;224:343–58. https://doi.org/10.1016/0022-2836(92)90999-z.

[47] Rabbitts TH, Forster A, Milstein CP. Human immunoglobulin heavy chain genes: evolutionary comparisons of Cμ, Cδ and Cγ genes and associated switch sequences. Nucleic Acids Research 1981;9:4509–24. https://doi.org/10.1093/nar/9.18.4509.

Figure 1. Multi-technique molecular diagnosis of the studied family A. Familial pedigree and STRs segregation analysis. Family tree was drawn based on the information provided by the family during the anamnesis. "del": deletion; "-": STR not analyzed; "dashed line square": alleles found to have 1-repeat retraction in III1. **B**. The upper figure shows a schema of the *DMD* cxSV found in III4. The lower figure depicts the MLPA results, the deletion is shown with a dashed-line rectangle and the duplication with a dotted-line rectangle. **C**. HUMARA electropherograms and XIP for II1 and II3. "Mock": Non-digested DNA; "HpaII": HpaII-digested DNA; "n": *AR* exon 1 STR alleles; "AUC": Area under the curve.

Figure 2. Molecular characterization of the cxSV in III4 A. The figure shows the SNP-array results for the *DMD* locus. The deletion and duplication breakpoints are marked with rectangles (lower panel) and zoomed in (upper panel) to show the SNPs reference sequence (rs). "BAF": B-allele frequency; "LRR": Log R ratio; "+": single-copy SNP; "-": null genotype; "++": double-copy SNP; "triangles": LR-PCR primers (AF1, AR1 and AR2). **B**. The figure shows the IGV screenshot for *DMD* WGS. The upper figure shows the sequencing depth-coverage for the cxSV *locus*, the absence of black bars denote a region that was not sequenced (deletion) and higher black bars represent the duplicated *loci*. This figure also shows the SNP analyzed in the array and the circles mark the cxSV breakpoints. The middle panel depicts a zoom in the breakpoint *loci*, and the chimeric reads are denoted with circles. The bottom panel, shows a detailed schema of the chimeric reads with their identification number, black spots delimit the sequence that mapped on each side of the breakpoint junction. "+":single-copy SNP; "-":null genotype; "++":double-copy SNP; "black triangles": PCR primers (DUP_F and DUP_R). **C**. Sanger sequencing of the breakpoint junctions from deletion (upper panel) and duplication (bottom panel). DNA sequences were aligned to *DMD* Reference Sequence NG_012232.1, GRCh38.p13. The deletion presented a 1 "T" microhomology (bolded), and the duplication showed an 18 bp insertion at the specific head-to-tail fusion (bolded).

Figure 3. Analysis of secondary structures at SV breakpoints The figure shows the DNA secondary structures predicted by Mfold. We used a 50 bp sequence of the reference genome (GRCh38.p13) surrounding the SV breakpoints (grey boxes) of the deletion (**A**) and the duplication (**B**). The sites of directed or inverted insertions of the duplication were also included in the analysis. "ΔG": Gibbs free energy [kcal/mol] associated with each secondary structure.

Figure 4. Schematic model of the suggested recombination mechanism leading to the deletion
A. Schema of *DMD* gene showing the orientation towards the telomere (Xp tel). **B**. The panel depicts the Paternal and Maternal Homologous X-chromosomes. "Black rectangles": maternal *DMD* exons; "White rectangles": paternal *DMD* exons; STR"n": where number indicate intronic location (introns 44, 45, 49 and 62); "STRs in bold": Maternal alleles; "STRs underlined": Paternal alleles; "Black zigzag lines": eventual crossover breakpoints; "Chevrons": Repetitive elements detected by Dfam algorithm ("Vertical lines": LINEs, "Dotted": LTRs, "Black": MIRs, "White": Alus, "Horizontal lines": others); "Up-pointing black triangle": Other relevant tandem repeats possibly involved in DNA ruptures as breakpoint stimulators [(TG)n and (T)n]. **C**. Unequal inter-chromosome recombination model. Schematic representation of a non-allelic non-homologous recombination event explaining both the origin of the deletion and the recombination between STR44 and STR62. **D**. Resulting III4´s chimeric X-chromosome structure, showing the deletion as the absence of STR45 and STR49.

Table 1: DNA motifs surrounding the breakpoints junctions

SV	Microhomologies at breakpoints	Repetitive elements		Non-B DNA		Secondary structures		Recombinogenic DNA motifs	
		5'	3'	5'	3'	5'	3'	5'	3'
Del	T	-	LTR	-	-	Yes	Yes	Jurka, SAR	Ig heavy chain switch
Dup	T, -, AG	LINE L1, -, LINE L1	-, LINE L1, -	-, -, -	-, -, cruciform motif	Yes, Yes, Yes	Yes, Yes, No	Jurka, Jurka, Jurka	Jurka, Jurka, Jurka

LTR: long terminal repeats; Jurka: hexanucleotide motifs targeted by the endonuclease/retro-transcriptase of mammalian retroposons; SAR: scaffold attachment region; LINE: long interspersed nuclear elements.