A MULTI-TECHNIQUE MOLECULAR ANALYSIS OF A DMD FAMILY WITH TWO INDEPENDENT MUTATIONAL EVENTS AND A MANIFESTING PREGNANT WOMAN. BIOINFORMATIC ANALYSIS OF MOLECULAR SCARS AT BREAKPOINT JUNCTIONS AND HYPOTHESIZATION OF THE UNDERLYING MOLECULAR MECHANISMS.

Leonela Luce[1], Miguel Abelleyro[2], Micaela Carcione[3], Chiara Mazzanti[3], Liliana Rossetti[2], Pamela Radic[2], Irene Szijan[3], Sebastián Menazzi [4], Liliana Francipane[4], Julián Nevado[5], Pablo Lapunzina[5], Carlos De Brasi[2], and Florencia Giliberto[1]

[1]Universidad de Buenos Aires. Facultad de Farmacia y Bioquímica. Departamento de Microbiología, Inmunología, Biotecnología y Genética, Cátedra de Genética, Laboratorio de Distrofinopatías.
[2]CONICET-Academia Nacional de Medicina. Instituto de Medicina Experimental (IMEX).
[3]Universidad de Buenos Aires. Facultad de Farmacia y Bioquímica. Departamento de Microbiología, Inmunología, Biotecnología y Genética, Cátedra de Genética, Laboratorio de Distrofinopatías.
[4]Universidad de Buenos Aires. Hospital de Clínicas "José de San Martín". División de Genética.
[5]Instituto de Genética Médica y Molecular (INGEMM)-IdiPAZ, Hospital Universitario La Paz, Universidad Autónoma

June 17, 2020

Abstract

Our work depicts a familial Duchenne muscular dystrophy case with a complex structural variant (cxSV) and a manifesting pregnant woman. Were our aims to provide molecular diagnosis and hypothesize mechanisms underlying the origin of the cxSV. We implemented a multi-technique approach including MLPA, STRs-segregation, AR-assay, SNP-array, WGS and a bioinformatic algorithm for identification of double strand breaks (DSB) stimulator motifs. We established the carrier status of the prenatal sample and explained its mother´s symptomatology by skewed X-chromosome inactivation. Furthermore, an ancestral familial ex38-43 duplication plus a de novo ex45-54 deletion was revealed in the proband, who carried the cxSV in a recombinant maternal X-chromosome. Characterization of cxSV´s breakpoints junction and its surrounding sequences allowed us to identify DSB stimulator motifs. The replication-dependent "Fork Stalling and Template Switching" mechanism was predicted to be the most likely scenario for the duplication´s origin. Whilst, the de novo deletion could arise from a germline event of inter-chromosome non-allelic recombination involving the "Non-Homologous End Joining" mechanism. The multi-technique strategy enabled precise diagnosis, accurate genetic assessment and widen the understanding of the molecular mechanisms involved in SVs' generation. Finally, the further comprehension of the occurrence of DMD variants, favors the development of new gene therapy strategies.

1

INTRODUCTION

Dystrophinopathies are X-linked recessive diseases caused by pathogenic variants in *DMD* gene (OMIM ID: 300377). Under this term are included different clinical features, covering a spectrum of muscle disease ranging from mild to severe that includes Duchenne Muscular Dystrophy (DMD), Becker Muscular Dystrophy (BMD) and DMD-associated Dilated Cardiomyopathy (DCM) (Brandsema & Darras, 2015; Darras et al., 2000). DMD is the most prevalent pediatric form of muscular dystrophy, with an incidence of 1:3500-5000 male births (Mendell et al., 2012). The disease has a *de novo* mutation rate of 33%, that is why there exist many families without previous history of the disease (Haldane, 2004). DMD is mainly generated by a complete absence of the dystrophin protein, which produces early muscle degeneration, leading to increase serum levels of Creatine Kinase (CK) (Emery, 1977).

Despite Dystrophinopathies are X-linked recessive diseases with a complete penetrance in males, in heterozygous females carriers the penetrance varies and may depend on X-chromosome inactivation patterns (XIP) (Darras et al., 2000). Pegoraro *et al,* have showed that more than 90% of heterozygous females with skewed XIP develop signs of muscular dystrophy (Pegoraro et al., 1995). This correlation was also observed by Giliberto *et al* and Viggiano *et al*(Giliberto et al., 2014; Viggiano et al., 2016, 2017).

DMD pathogenic variants comprise mainly Structural Variants (SVs), such as deletions (68%) or duplications (11%) of one or more exons, and small mutations (20%) (Aartsma-Rus et al., 2016). Precise deleterious variant identification and accurate molecular diagnosis are crucial for Dystrophinopathy patients not only to access to the specific and optimal standard of care and to achieve family planning, but also provides information on eligibility for mutation-specific treatments.

SVs are defined by the novel sequence generated at the breakpoints junctions. Different mutational mechanisms have been hypothesized to explain their origin. Among them can be highlighted: 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 which can take place in meiosis or in 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 (Hastings et al., 2009; Lee et al., 2007; McEachern & Haber, 2006; McVey & Lee, 2008; Moore & Haber, 1996; Quinlan & Hall, 2012; Szostak et al., 1983).

The analysis of the surrounding breakpoints sequences, i.e. screening of DNA instability and DSB stimulators motifs, provide hints about which molecular mechanisms may be involved in the origin of SVs. Several and heterogenous motifs have been described in literature, among them: 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 (Brouha et al., 2003; Deininger & Batzer, 1999; Kazazian, 2004; Rüdiger et al., 1995); Non-B DNA, such as tetraplex DNA, cruciform DNA, bent DNA, Z-DNA, and all sort of secondary structures (Bacolla & Wells, 2004; Wang & Vasquez, 2014); Short Tandem Repeats (STRs), which were found enriched at DSBs and in *DMD* intron breakpoint hotspots (Luce et al., 2016; Zavodna et al., 2018); and finally, Alu/LINE specific retro-transposition target sequences (Been et al., 1984; Jurka, 1997; Spitzner & Muller, 1988; Weaver & DePamphilis, 1982).

Here, we present a familial case of DMD with a symptomatic pregnant woman and a thorough analysis of the complex SVs (cxSVs) found. Were our aims to carry out precise prenatal diagnosis and to hypothesize the molecular mechanisms underlying the origin of the SV, implementing a multi-technique molecular algorithm.

MATERIALS AND METHODS

Patients

A DMD familial case was referred to our laboratory. The index case was a 35-year-old symptomatic pregnant woman (II1) who was pursuing a prenatal diagnosis. She has a deceased affected brother (II2), diagnosed with DMD by muscle biopsy. Another key relative was a 17-year-old nephew (III4), who was previously diagnosed by multiplex-PCR with a *DMD* 45-54 exons deletion. Pedigree is shown in figure 1A.

The pregnant woman referred difficulties at walking and climbing stairs since she was 21-year-old, a muscle biopsy compatible with DMD, 3,100IU/l CK levels and a normal female karyotype (46,XX).

Samples

Peripheral blood leukocytes from the symptomatic pregnant woman (II1), the affected nephew (III4) and his mother (II3), and chorionic villus sampling (CVS) from the fetus (III1) were referred to our laboratory (Figure 1A).

Whole blood was drawn by venipuncture with 5% ethylene-diamine tetraacetic acid (EDTA) as anticoagulant. Genomic DNA (gDNA) from blood samples was isolated using the cetyl-trimethyl-ammonium bromide (CTAB) method (Murray & Thompson, 1980). For the CVS, gDNA was isolated by QIAGEN DNeasy Blood and tissue kit [Redwood City, California (*www.qiagen.com*)]. DNA concentration and quality were measured by absorbance at 260nm and by the ratios A260nm/A280nm and A260nm/A230nm, respectively.

The protocol was approved by the institutional ethics committee. Informed consent was obtained for all study subjects prior to the molecular studies.

Haplotype Assay

STRs-(CA)n analysis was designed in order to perform intrafamilial deletion segregation and discard CVS maternal blood contamination. We amplified up to 5 *DMD* intronic microsatelites in family members [STRs: DYSII, 44, 45, 49 and 62] (Beggs & Kunkel, 1990; Clemens et al., 1991; King et al., 1995). The amplified microsatellites were labeled using 6-FAM-primers. PCR was performed as previously described elsewhere, with minor modifications (Luce et al., 2014). Primer sequences were obtained from the Leiden Muscular Dystrophy site [www.dmd.nl]. All PCR reactions were performed in a thermal cycler [Veriti; Applied Biosystems, Foster City, California]. PCR products were analyzed using a fragment analyzer sequencer [ABI 3730XL; Applied Biosystems, Foster City, California]. Data analysis was performed using PeakScanner software [Applied Biosystems, Foster City, California].

Multiplex Ligation-dependent Probe Amplification (MLPA)

MLPA kit for *DMD* (Salsas PO34-PO35) was used to find deletions/duplications (Gatta et al., 2005; Janssen et al., 2005; Schwartz & Dunø, 2004). 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] and 500Liz as internal size standard. Data analysis was performed using Coffalyser [MRC-Holland, Amsterdam, The Netherlands] and GeneMarker V2.2.0 [Softgenetics, State College, Pennsylvania] softwares. Wildtype, deleted, and duplicated controls were included.

Human Androgen Receptor X-Chromosome Inactivation Assay (AR-assay)

The X-chromosome inactivation pattern (XIP) was performed on II1 and II3 by a methylation-based PCR of the androgen receptor (*AR* Xq12) first exon, as described by Allen *et al*(Allen et al., 1992).*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].

High density SNP-array

DNAs were quantified using PicoGreen [Invitrogen Corporation, Carlsbad, United States]. A genome-wide scan of 850,000 tag SNPs was conducted on III4, using Illumina CytoSNP-850k BeadChip according to manufacturer's specifications [Illumina, San Diego, United States]. GenCall scores <0.15 at any locus were considered "no-calls". Image data were analyzed using the Chromosome Viewer tool contained in Genome Studio [Illumina, San Diego, United States]. The metric used was the logR ratio which is the log (base 2) ratio of the observed normalized R-value for a SNP divided by the expected normalized R-value (under manufacturer's specifications). In addition, an allele frequency analysis was applied for all SNPs. Genomic positions were based upon GRCh37.

3

Whole genome sequencing (WGS) analysis

WGS was performed in III4 using a HiSeq instrument [Illumina, San Diego, United States] by Macrogen services [Republic of Korea]. Bioinformatics analysis included a quality control with FASTX-toolkit (v.0.0.13.2), reads mapping 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).

Deletion Long Range-PCR (LR-PCR)

LR-PCR primers were designed according to inferred 5' and 3' breakpoints from SNP-array (Supplementary table 1). Amplification products ranged from 2.0 to 3.5kb. Reactions were performed following the enzyme's manufacturer recommendations [KAPA Biosystem, Wilmington, United States] in a thermal cycler [Hangzhou Bioer Technology, China]. PCR products were analyzed by Sanger sequencing.

PCR amplification of duplication's breakpoints

The specific PCR amplification for the tandem duplication 5' and 3' breakpoints were based on the hypothesis of head-to-tail segmental fusion and adjusted by WGS results (Supplementary table 1). Reactions were performed following the enzyme's manufacturer recommendations [Promega, Madison, United States] in a thermal cycler [Biometra, Germany]. Products were characterized by Sanger Sequencing.

Deletion and duplication breakpoints bioinformatic analysis

DNA intervals ranging from 10bp to 50bp centered on each 5' and 3' breakpoint of deletion and duplication were *in silico* screened for microhomologies, repetitive elements, non-B DNA, secondary structures and recombinogenic DNA motifs. These elements constitute a heterogeneous group of sequences with different lengths (3-18bp), 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 (Abelleyro et al., 2020). For this analysis, DSB stimulation motifs that showed significant Expected values (E-values <0.05) in random points from the referred study were considered (Abelleyro et al., 2020). 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/retrotranscriptase of mammalian retroposons (Jurka motifs) (Jurka, 1997).

RESULTS

Haplotype Assay

STRs segregation study resulted informative, as 4/5 of the analyzed *loci* were heterozygous. Conformed haplotypes are depicted in figure 1A. Haplotype analysis not only allowed us to exclude CVS's maternal blood contamination, as the CVS presented only 1 maternal X-chromosome, but also predicted a female gender given that two X-chromosomes were detected. Noteworthy, a 1-step retraction of STR49 was also observed in the fetus. On the other hand, we could bring out the previously reported deletion in III4 by the absence of amplification for STR45 and STR49. Granted that II1, II3 and III1 resulted heterozygous for STR49, we could predict that they didn't carry III4's deletion, however, they share a fraction of the at-risk haplotype (STRDYSII and STR44). Moreover, sisters II1 and II3 showed a different paternal haplotype.

4

Finally, another important result was the observation of a recombinant X-chromosome, between STR44 and STR45, in III4.

MLPA

MLPA revealed a cxSV in III4. Not only did he carry the already known deletion of exons 45-54 but also a duplication of exons 38-43 (Figure 1B). Nevertheless, II1, II3 and III1 only inherited the duplication involving exons 38-43 (data not shown). These results are in concordance with the ones observed in the haplotype assay.

AR-ASSAY

Study revealed a skewed XIP of 96% in II1 (alleles *21* /25,*active allele*), and random XIP of 65% in II3 (alleles 23/20). Sisters II1 and II3 share none of the *AR* alleles (Figure 1C).

Deletion's molecular characterization

SNP-array analysis on III4 detected that the deletion spans about 450kb. Also, it allowed us to delimit the 5' deletion breakpoint within an interval defined by SNPs rs1950112 (NC_00023.11:g.32095444) and rs1795577 (NC_00023.11:g.32092321), while the 3' breakpoint by SNPs rs2030002 (NC_00023.11:g.31646698) and rs5972426 (NC_00023.11:g.31646233). Given that distances between these SNPs were short, we could design a LR-PCR to sequence the breakpoint junction (Figure 2A). Combination of primers AF1 and AR1 did not amplified, suggesting that AR1 mapped inside the deletion. Whilst AF1 and AR2 amplified a product of approximately 2,435bp, delimiting the deletion to 446,477bp (NM_004006.3:c.6438+123812_8027+11362del). This deletion breakpoint junction was confirmed by the several aligned chimeric reads (i.e., reads that did not map entirely on the reference sequence, such as ID6762, ID9251, ID26052) obtained by WGS (Figure 2B, Supplementary Figure 1A). Lastly, characterization of the deletion breakpoint junction showed a single base "T" of microhomology between introns 44 and 54 (Figure 2C, Table 1).

Duplication's molecular characterization

SNP-array from III4 also detected the duplication previously observed by MLPA, which spans ~124kb. We were able to delimitate the 5' breakpoint within an interval defined by SNPs rs12856332 (NC_00023.11:g.32367273) and rs1801187 (NC_00023.11:g.32362879), while the 3' breakpoint by SNPs rs111931446 (NC_00023.11:g.32229089) and rs143786489 (NC_00023.11:g.32227327). As distances between these SNPs were too long (~4kb and ~2kb, respectively), SNP-array results alone did not offer enough data to set up a specific PCR amplification system even under the head-to-tail fusion hypothesis (Figure 2A).

WGS results were essential to chase the unique singularity of the duplication's breakpoint junction. The identification of chimeric reads, formed by sequences of introns 37 and 43, mapping in the limits of double coverage region (e.g. ID22075, ID17613, ID18803, ID15702) were consistent with a head-to-tail tandem duplication (Figure 2B, Supplementary Figure 1B). Noteworthy, we could find out 2 contradictions between SNP-array and WGS, as rs1801187 and rs143786489 which were supposed to be in double and single copy, respectively, by the array turned out to be in single and double copy.

The alignment of WGS chimeric overlapping reads allowed the determination of the breakpoint sequence. This information permitted the design of a duplication specific head-to-tail PCR of 366bp, in order to confirm the duplication breakpoint characterization by Sanger sequencing (Figure 2C). The size of the duplicated region was defined by 131,284bp (NM_004006.3:c.6291-5371_6291-5370ins[TAAAATGCAATTTCATTT;5326-5188_6291-5370]). Finally, the characterization of the duplication breakpoint junction showed a complex rearrangement, formed by a 7bp inverted insertion, followed by an 11bp direct insertion both with intron 43 sequence identity. This suggest 3 events of template switching (TS) with microhomology of 1bp "T" at first TS and a microhomology of 2bp "TC" at third TS (Figure 2C, Table 1).

Bioinformatics analysis of SV breakpoints

The performed analysis showed that all breakpoints had at least a significant DBS stimulator motif. The contingency of DNA motifs found within each interval around the breakpoints from the deletion and duplication are shown in Table 1.

Repetitive elements were found in 4/8 breakpoint junctions of the SVs, among them LTR (n=1), LINE L1 (n=3) (Table 1). On the other hand, only 1/8 breakpoints presented a non-B DNA structure. A cruciform DNA motif was found in the last 3' breakpoint of the duplication (Table 1). Regarding the analysis by mfold algorithm, every breakpoint junction showed to adopt a secondary DNA structure. However, 7/8 were counted as positives, those in which the breakpoint co-localize with the secondary structure itself (Table 1, Figure 3). The most stable identified structures presented heterogeneous [?]G values, ranging from -5.01 to -0.27 with an average value of -1.77kcal (Figure 3).

Based on the hypothesis for the deletion's origin of a non-allelic non-homologous inter-chromosomal recombination assisted by events of ectopic synapsis between repeats, we broaden the analysis of the breakpoint junctions with Dfam to a 5kb interval. The Dfam algorithm detected, at least, 8 partial- and full-length repetitive elements (LINEs, Alus, LTRs and MIRs, etc.) that may be involved in the formation of the ectopic synapsis (Figure 4). Furthermore, 2 tandem (TG)n and (T)n repeats were found in the analyzed interval, which may act as DNA breaks stimulators (Figure 4).

DISCUSSION

DMD is one of the most prevalent among the rare diseases (Mendell et al., 2012). It is caused by sequence variants in *DMD* , the largest gene of the human genome, which is located in a genomic region with high rates of recombination. These characteristics make *DMD* highly susceptible to mutation (Nguyen & Yokota, 2019). The current work presents a familial case of DMD with a cxSV, a multi-technique and bioinformatic approach was implemented in pursuit of 2 major aims. The first one focus on the molecular diagnosis and genetic assessment of the family, including manifesting carrier and prenatal studies. While the second one, centered on finding possible explanations of the origin of SVs in *DMD* .

The design of the prenatal diagnosis molecular strategy was based on the previously detected *DMD* deletion of exons 45-54 in the affected alive male (III4). On the one hand, we performed a CVS maternal cell contamination study to determine the usefulness of the fetal material. On the other hand, we decided to evaluate the presence of the deletion by 2 different means, MLPA and the co-segregation of the deletion by STRs. Therefore, 2 STRs mapping within the deletion and 2 flanking it were selected for the haplotype study.

A priori, we were expecting that II1 and II3 shared III4's deletion, so it was a surprise when studies revealed the existence of 2 SVs in the family. A duplication of exons 38-43, shared by all family members, probing to be the ancestral familial causing mutation. Also, the well-known deletion, found only in III4, suggesting being the second mutational event i.e. a *de novo* alteration. Therefore, this case works as an example of the importance of retesting historic patients diagnosed by multiplex-PCR, where deletion's boundaries might not have been precisely determined or other SVs could have been missed out, affecting eligibility and effectiveness of mutation-dependent therapies.

We would like to hallmark the usefulness of the old-fashioned STRs segregation studies as a tool to identify homologous recombination and STRs retraction/expansion events, determine parental origin of the haplotypes, corroborate sample's gender and discard CVS maternal cell contamination. In this case, STRs revealed a recombination between STR44 and STR62 in III4, a retraction of STR49 in III1, the existence of 2 different paternal X-chromosomes in II1 and II3, an estimation of female gender in the fetus and the exclusion of CVS contamination.

Regarding the AR-Assay, the skewed XIP observed in II1 suggests that the fully active X-chromosome should be the maternal one, carrying the duplication, giving the presence of DMD symptomatology. On the contrary, as expected, her asymptomatic sister (II3) had a random inactivation. Because of the STRs and MLPA results, II1 and II3 were already expected to present different paternal X-chromosomes but to

6

share the maternal one, as both carried the duplication. Remarkably, they did not share any AR allele, corroborating the double paternity and suggesting another recombination event that separated the DMD causative mutation from the AR allele.

Although, AR-assay has been recently validated on amniotic fluid for DMD symptomatology prediction, we could not test the XIP in the fetus, as their parents did not agree on performing a second puncture (He et al., 2019). There are conflicting reports in the literature about the heritability of the XIP (Renault et al., 2007; Viggiano et al., 2017). Therefore, future XIP studies may be needed in case any clinical symptoms arise in III1.

Several authors reported the occurrence of non-contiguous rearrangements, within the same DMD allele, with frequencies up to 2% (Kerr et al., 2013). In our studied cohort, we estimated a cxSV rate of 1.4% (6/437 DMD patients), encompassing deletions-duplications, non-contiguous duplications and large deletion plus a 20pb insertion. The present case offered an extra advantage, the possibility of establishing a mutational timeline thanks to familial segregation study of the detected cxSV.

Molecular diagnosis encompasses the study of index cases, usually lacking a complete familial analysis. The obtained molecular alterations have a "photo effect", like a picture of the resulting rearrangement at a specific moment careless of timeline determination. This lack of information is more obvious when dealing with cxSVs, where is impossible to differentiate between 2 concurrent alterations and 2 different mutational events that take place in different times/generations. Studying chronologically these cxSVs could be useful to unravel if the first alteration acts as a predisposing factor to the latter.

Focusing on our second aim, centered on finding the molecular mechanisms underlying the origin of the observed SVs, we decided to precisely determine their breakpoints. For this purpose, we used several practical approaches, including SNP-array, WGS and customized PCRs. Followed by a bioinformatic approach, screening the surroundings of the SVs' breakpoints for the identification of DNA DSB stimulator motifs, which have been found statistically significantly more frequent at breakpoint junctions than expected by chance according to Abelleyro *et al*(2020) (Abelleyro et al., 2020).

While deletion breakpoints were neatly and accurately identified by SNP-array and LR-PCR, the characterization of duplication breakpoints was particularly challenging, as the SNP-array could not certainly delimitate the single/double-doses transitions. It was only after WGS analysis that the identification of the reads mapping on the tandem duplication breakpoint junction permitted its precise delimitation. Therefore, these results remark the importance of revalidating the SV's boundaries detected by SNP-array.

Regarding the tandem segmental duplication, this event may be adequately explained either by MMBIR or FoSTeS models, both involving *de novo* DNA synthesis. However, considering the principle of maximum parsimony and the complexity of the event, characterized by at least 3 strand invasions preceded by replication fork collapses and facilitated by microhomologies, supports a FoSTeS mechanism over MMBIR (Figure 5A). In addition, the necessary DNA strand break collapses may have been stimulated by the presence of Jurka hexa-nucleotides, secondary structures as well as repetitive elements at relevant locations (Vissers et al., 2009).

Concerning the deletion, it seems to occur *de novo* concurrently with a meiotic recombination event switching STRs alleles between maternal homologous, given the coincident X-chromosome location of both molecular events. The most likely scenario suggests that these two concomitant molecular events may actually represent a single inter-chromosomal recombination event, which could have taken place either in II3's oocyte meiosis or in II3's oogonia mitosis, making the deletion inheritable. The above mentioned evidence indicates that II3 carries the DMD duplication, but does not have the deletion in peripheral blood, suggesting that the complex series of events took place in her germline resulting in a chimeric X-chromosome with STRs alleles from both progenitors (I1 and I2) and both SVs in phase. Furthermore, our recombination hypothesis relies on the presence of repetitive elements relatively near, both, 5' and 3' deletion breakpoints, supporting a putative inter-chromosome non-allelic pairing structure or ectopic synapsis suitable for a localized recombination (Figure 4) (Liu et al., 2012).

7

At the molecular level, the breakpoint junctions of the deletion point out the hypothesis of a classical NHEJ, given that the event does not present any molecular characteristic that could imply an association with a DNA replication dependent mechanism. We detected just a single-thymine of microhomology, which does not provide sufficient evidence to sub-classify the event as MMEJ (Figure 5B). Moreover, supporting the classical NHEJ model DNA DSB stimulating motifs were found at both deletion breakpoints, such as Jurka hexa-nucleotides and SAR on the 5' breakpoint and Ig heavy chain switch region on the 3' deletion breakpoint (Gale et al., 1992; Jurka, 1997; Rabbitts et al., 1981). Other possible stimulators for DBS, possibly exerting its action, could be LTRs, STRs and highly stable secondary structures, which have also been detected at both deletion breakpoints. Furthermore, 2 of these motifs might be related, as long STRs have the ability to form non-B-DNA structures.

Resuming the chronology of the molecular events that gave rise to the cxSV, the duplication could have favored the approach of the regions involved in the deletion and stabilize an unequal pairing between the duplicated X-chromosome and its unduplicated counterpart. This theory supports both previously mentioned hypothesis for the origin of the deletion. On the 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.

On the other hand, we want to highlight the important role that secondary structures must be exerting in the origin of SVs, given that we found them in 7/8 breakpoints. Its implication for DSB formation is suggested by the fact that if we randomly select 50pb throughout the genome it is very likely that they form secondary structures, but if we simulate random breakpoints and analyze 25bp at each side of it, the chances of finding secondary structures involving them are greatly reduced (Abelleyro et al., 2020).

In conclusion, the strategy implemented in the present work allowed us to provide an accurate diagnosis and genetic assessment, enabling early diagnosis and the selection of the appropriate standard of care and mutation-specific treatments. Finally, the thorough characterization of the SV's breakpoints and the analysis of their molecular scars not only widen the understanding of the molecular mechanisms involved in their generation, but also may help the development of new therapeutic approaches.

ACKNOWLEDGEMENTS

CONFLICT OF INTERESTS

The authors declare no conflict of interests.

Data Availability Statement

Research data are not shared.

REFERENCES

Aartsma-Rus, A., Ginjaar, I. B., & Bushby, K. (2016). The importance of genetic diagnosis for Duchenne muscular dystrophy. In *Journal of Medical Genetics* (Vol. 53, Issue 3, pp. 145–151). https://doi.org/10.1136/jmedgenet-2015-103387

Abelleyro, M. M., Radic, C. P., Marchione, V. D., Waisman, K., Tetzlaff, T., Neme, D., Rossetti, L. C., & De Brasi, C. D. (2020). 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. *Human Mutation*, *41*(4), 825–836.

Allen, R. C., Zoghbi, H. Y., Moseley, A. B., Rosenblatt, H. M., & Belmont, J. W. (1992). Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation.*American Journal of Human Genetics*, *51*(6), 1229–1239.

Bacolla, A., & Wells, R. D. (2004). Non-B DNA Conformations, Genomic Rearrangements, and Human Disease. In *Journal of Biological Chemistry* (Vol. 279, Issue 46, pp. 47411–47414). https://doi.org/10.1074/jbc.r400028200

Been, M. D., Burgess, R. R., & Champoux, J. J. (1984). Nucleotide sequence preference at rat liver and wheat germ type 1 DNA topoisomerase breakage sites in duplex SV40 DNA. *Nucleic Acids Research*, *12*(7), 3097–3114.

Beggs, A. H., & Kunkel, L. M. (1990). A polymorphic CACA repeat in the 3' untranslated region of dystrophin. In *Nucleic Acids Research* (Vol. 18, Issue 7, pp. 1931–1931). https://doi.org/10.1093/nar/18.7.1931

Brandsema, J. F., & Darras, B. T. (2015). Dystrophinopathies. *Seminars in Neurology*, *35*(4), 369–384.

Brouha, B., Schustak, J., Badge, R. M., Lutz-Prigge, S., Farley, A. H., Moran, J. V., & Kazazian, H. H., Jr. (2003). Hot L1s account for the bulk of retrotransposition in the human population. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(9), 5280–5285.

Clemens, P. R., Fenwick, R. G., Chamberlain, J. S., Gibbs, R. A., de Andrade, M., Chakraborty, R., & Caskey, C. T. (1991). Carrier detection and prenatal diagnosis in Duchenne and Becker muscular dystrophy families, using dinucleotide repeat polymorphisms. *American Journal of Human Genetics*, *49*(5), 951–960.

Darras, B. T., Urion, D. K., & Ghosh, P. S. (2000). Dystrophinopathies. In M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. Stephens, & A. Amemiya (Eds.), *GeneReviews*. University of Washington, Seattle.

Deininger, P. L., & Batzer, M. A. (1999). Alu repeats and human disease. *Molecular Genetics and Metabolism*, *67*(3), 183–193.

Emery, A. E. H. (1977). Muscle histology and creatine kinase levels in the foetus in Duchenne muscular dystrophy. In *Nature* (Vol. 266, Issue 5601, pp. 472–473). https://doi.org/10.1038/266472a0

Gale, J. M., Tobey, R. A., & D'Anna, J. A. (1992). Localization and DNA sequence of a replication origin in the rhodopsin gene locus of Chinese hamster cells. In *Journal of Molecular Biology* (Vol. 224, Issue 2, pp. 343–358). https://doi.org/10.1016/0022-2836(92)90999-z

Gatta, V., Scarciolla, O., Gaspari, A. R., Palka, C., De Angelis, M. V., Di Muzio, A., Guanciali-Franchi, P., Calabrese, G., Uncini, A., & Stuppia, L. (2005). Identification of deletions and duplications of the DMD gene in affected males and carrier females by multiple ligation probe amplification (MLPA). In *Human Genetics* (Vol. 117, Issue 1, pp. 92–98). https://doi.org/10.1007/s00439-005-1270-7

Giliberto, F., Radic, C. P., Luce, L., Ferreiro, V., de Brasi, C., & Szijan, I. (2014). Symptomatic female carriers of Duchenne muscular dystrophy (DMD): genetic and clinical characterization. *Journal of the Neurological Sciences*, *336*(1-2), 36–41.

Haldane, J. B. S. (2004). The rate of spontaneous mutation of a human gene. In *Journal of Genetics* (Vol. 83, Issue 3, pp. 235–244). https://doi.org/10.1007/bf02717892

Hastings, P. J., Ira, G., & Lupski, J. R. (2009). A Microhomology-Mediated Break-Induced Replication Model for the Origin of Human Copy Number Variation. In *PLoS Genetics* (Vol. 5, Issue 1, p. e1000327). https://doi.org/10.1371/journal.pgen.1000327

He, W.-B., Du, J., Xie, P.-Y., Zhou, S., Zhang, Y.-X., Lu, G.-X., Lin, G., Li, W., & Tan, Y.-Q. (2019). X-chromosome inactivation pattern of amniocytes predicts the risk of dystrophinopathy in fetal carriers of DMD mutations. *Prenatal Diagnosis*, *39*(8), 603–608.

Janssen, B., Hartmann, C., Scholz, V., Jauch, A., & Zschocke, J. (2005). MLPA analysis for the detection of deletions, duplications and complex rearrangements in the dystrophin gene: potential and pitfalls. In *Neurogenetics* (Vol. 6, Issue 1, pp. 29–35). https://doi.org/10.1007/s10048-004-0204-1

Jurka, J. (1997). Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(5), 1872–1877.

Kazazian, H. H., Jr. (2004). Mobile elements: drivers of genome evolution. *Science*,*303*(5664), 1626–1632.

Kerr, R., Robinson, C., Essop, F. B., & Krause, A. (2013). Genetic testing for Duchenne/Becker muscular dystrophy in Johannesburg, South Africa. *South African Medical Journal = Suid-Afrikaanse Tydskrif Vir Geneeskunde*,*103*(12 Suppl 1), 999–1004.

King, S. C., Roche, A. L., Rita Passos-Bueno, M., Takata, R., Zatz, M., Cockburn, D. J., Seller, A., Stapleton, P. M., & Love, D. R. (1995). Molecular characterization of further dystrophin gene microsatellites. In *Molecular and Cellular Probes* (Vol. 9, Issue 5, pp. 361–370). https://doi.org/10.1016/s0890-8508(95)91700-4

Lee, J. A., Carvalho, C. M. B., & Lupski, J. R. (2007). A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders.*Cell*, *131*(7), 1235–1247.

Liu, P., Carvalho, C. M. B., Hastings, P. J., & Lupski, J. R. (2012). Mechanisms for recurrent and complex human genomic rearrangements. In *Current Opinion in Genetics & Development* (Vol. 22, Issue 3, pp. 211–220). https://doi.org/10.1016/j.gde.2012.02.012

Luce, L. N., Dalamon, V., Ferrer, M., Parma, D., Szijan, I., & Giliberto, F. (2016). MLPA analysis of an Argentine cohort of patients with dystrophinopathy: Association of intron breakpoints hot spots with STR abundance in DMD gene. *Journal of the Neurological Sciences*, *365*, 22–30.

Luce, L. N., Ottaviani, D., Ferrer, M., Szijan, I., Cotignola, J., & Giliberto, F. (2014). Molecular diagnosis of dystrophinopathies using a multi-technique analysis algorithm. In *Muscle & Nerve* (Vol. 49, Issue 2, pp. 249–256). https://doi.org/10.1002/mus.23906

McEachern, M. J., & Haber, J. E. (2006). Break-induced replication and recombinational telomere elongation in yeast. *Annual Review of Biochemistry*, *75*, 111–135.

McVey, M., & Lee, S. E. (2008). MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. In *Trends in Genetics* (Vol. 24, Issue 11, pp. 529–538). https://doi.org/10.1016/j.tig.2008.08.007

Mendell, J. R., Shilling, C., Leslie, N. D., Flanigan, K. M., al-Dahhak, R., Gastier-Foster, J., Kneile, K., Dunn, D. M., Duval, B., Aoyagi, A., Hamil, C., Mahmoud, M., Roush, K., Bird, L., Rankin, C., Lilly, H., Street, N., Chandrasekar, R., & Weiss, R. B. (2012). Evidence-based path to newborn screening for Duchenne muscular dystrophy. *Annals of Neurology*, *71*(3), 304–313.

Moore, J. K., & Haber, J. E. (1996). Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. In *Molecular and Cellular Biology* (Vol. 16, Issue 5, pp. 2164–2173). https://doi.org/10.1128/mcb.16.5.2164

Murray, M. G., & Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. In*Nucleic Acids Research* (Vol. 8, Issue 19, pp. 4321–4326). https://doi.org/10.1093/nar/8.19.4321

Nguyen, Q., & Yokota, T. (2019). Antisense oligonucleotides for the treatment of cardiomyopathy in Duchenne muscular dystrophy. *American Journal of Translational Research*, *11*(3), 1202–1218.

Pegoraro, E., Schimke, R. N., Garcia, C., Stern, H., Cadaldini, M., Angelini, C., Barbosa, E., Carroll, J., Marks, W. A., Neville, H. E., Marks, H., Appleton, S., Toriello, H., Wessel, H. B., Donnelly, J., Bernes, S. M., Taber, J. W., Weiss, L., & Hoffman, E. P. (1995). Genetic and biochemical normalization in female carriers of Duchenne muscular dystrophy: evidence for failure of dystrophin production in dystrophin-competent myonuclei. *Neurology*, *45*(4), 677–690.

Quinlan, A. R., & Hall, I. M. (2012). Characterizing complex structural variation in germline and somatic genomes. *Trends in Genetics: TIG*, *28*(1), 43–53.

Rabbitts, T. H., Forster, A., & Milstein, C. P. (1981). Human immunoglobulin heavy chain genes: evolutionary comparisons of Cμ, Cδ and Cγ genes and associated switch sequences. In *Nucleic Acids Research* (Vol. 9, Issue 18, pp. 4509–4524). https://doi.org/10.1093/nar/9.18.4509

Renault, N. K., Dyack, S., Dobson, M. J., Costa, T., Lam, W. L., & Greer, W. L. (2007). Heritable skewed X-chromosome inactivation leads to haemophilia A expression in heterozygous females. *European Journal of Human Genetics: EJHG*, *15*(6), 628–637.

Rüdiger, N. S., Gregersen, N., & Kielland-Brandt, M. C. (1995). One short well conserved region ofAlu-sequences is involved in human gene rearrangements and has homology with prokaryoticchi. In *Nucleic Acids Research* (Vol. 23, Issue 2, pp. 256–260). https://doi.org/10.1093/nar/23.2.256

Schwartz, M., & Dunø, M. (2004). Multiplex ligation-dependent probe amplification is superior for detecting deletions/duplications in Duchenne muscular dystrophy. In*Clinical Genetics* (Vol. 67, Issue 2, pp. 189–191). https://doi.org/10.1111/j.1399-0004.2004.00382.x

Spitzner, J. R., & Muller, M. T. (1988). A consensus sequence for cleavage by vertebrate DNA topoisomerase II. *Nucleic Acids Research*, *16*(12), 5533–5556.

Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., & Stahl, F. W. (1983). The double-strand-break repair model for recombination. *Cell*, *33*(1), 25–35.

Viggiano, E., Ergoli, M., Picillo, E., & Politano, L. (2016). Determining the role of skewed X-chromosome inactivation in developing muscle symptoms in carriers of Duchenne muscular dystrophy. *Human Genetics*, *135*(7), 685–698.

Viggiano, E., Picillo, E., Ergoli, M., Cirillo, A., Del Gaudio, S., & Politano, L. (2017). Skewed X-chromosome inactivation plays a crucial role in the onset of symptoms in carriers of Becker muscular dystrophy. In *The Journal of Gene Medicine* (Vol. 19, Issue 4, p. e2952). https://doi.org/10.1002/jgm.2952

Vissers, L. E. L. M., Lisenka E L, Bhatt, S. S., Janssen, I. M., Xia, Z., Lalani, S. R., Pfundt, R., Derwinska, K., de Vries, B. B. A., Gilissen, C., Hoischen, A., Nesteruk, M., Wisniowiecka-Kowalnik, B., Smyk, M., Brunner, H. G., Cheung, S. W., van Kessel, A. G., Veltman, J. A., & Stankiewicz, P. (2009). Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genomic architecture. In*Human Molecular Genetics* (Vol. 18, Issue 19, pp. 3579–3593). https://doi.org/10.1093/hmg/ddp306

Wang, G., & Vasquez, K. M. (2014). Impact of alternative DNA structures on DNA damage, DNA repair, and genetic instability. In *DNA Repair* (Vol. 19, pp. 143–151). https://doi.org/10.1016/j.dnarep.2014.03.017

Weaver, D. T., & DePamphilis, M. L. (1982). Specific sequences in native DNA that arrest synthesis by DNA polymerase alpha. *The Journal of Biological Chemistry*, *257*(4), 2075–2086.

Zavodna, M., Bagshaw, A., Brauning, R., & Gemmell, N. J. (2018). The effects of transcription and recombination on mutational dynamics of short tandem repeats. In*Nucleic Acids Research* (Vol. 46, Issue 3, pp. 1321–1330). https://doi.org/10.1093/nar/gkx1253

TABLES

Table 1: DNA motifs surrounding the breakpoints junctions

SV
Del
Dup

LTR: long terminal repeats; Jurka: hexanucleotide motifs targeted by the endonuclease/retro-transcriptase of mammalian re

11

FIGURE LEGENDS

Figure 1. Molecular diagnostic studies. A . Familial pedigree and STRs segregation analysis. "del": deletion; "-": STR not analyzed; "dotted squares": alleles implicated in STR49 retraction in III1. **B.** Upper figure, schematization of the *DMD* cxSV from III4. Bottom figure, MLPA result highlighting the deletion with a red dotted rectangle and the duplication with a blue full-line rectangle. **C** . AR-assay electropherograms and XIP for II11 and II3. "Mock": Non-digested DNA; "HpaII": HpaII-digested DNA; "n":*AR* exon 1 STR allele; "AUC": Area Under Curve.

Figure 2. Characterization of the cxSV in III4. A . SNP-array analysis. Deletion and duplication breakpoints are marked by rectangles and zoomed in to show the SNPs reference sequence (rs) from 5' and 3' interval breakpoints. "BAF": B-allele frequency; "LRR": Log R ratio (LRR); "+": SNP present at normal dosage; "-": SNP with null dosage; "++": SNP with increased dosage; "black triangles": LR-PCR primers (AF1, AR1 and AR2) location for deletion breakpoints delimitation.**B.** WGS IGV analysis. Upper figure, circles on the coverage track (blue histogram) indicate regions of interest encompassing the SVs breakpoints, where the deletion is characterized by null dosage (no bars) and the duplication by an increased dosage (higher bars). Relevant SNP-array SNPs are also depicted on the coverage track. Insets show a zoom in of the breakpoint area, inner circles highlight key chimeric reads presenting mismatches with the Reference Genome. Bottom figure, schematization of the key chimeric reads with their identification number (left, deletion chimeric reads; right, duplication chimeric reads), black spots delimit the portion of the read that maped on each side of the breakpoint junction. "+": SNP present at normal dosage; "-": SNP with null dosage; "++": SNP with increased dosage; "black triangles": PCR primers (DUP_F and DUP_R) location for duplication breakpoints delimitation. **C** . Sanger sequencing of the breakpoints junctions from deletion (upper figure) and duplication (bottom figure). Obtained DNA sequences are aligned with *DMD*Reference Sequence (NG_012232.1, GRCh38.p13). The deletion present a 1 bp ("T") of microhomology highlighted in boldface. The duplication shows a 18 bp insertion at the specific head-to-tail fusion marked in boldface.

Figure 3. Secondary structure analysis of SVs breakpoints. A 50 bp interval of the reference genome (GRCh38.p13) centered on each 5' and 3' breakpoints (blue box) of the deletion (**A**) and the duplication (**B**) were analyzed by Mfold. Also, the sites of directed or inverted insertions of the duplication were included in the analysis. "bkp": breakpoint; "[?]G": Gibbs free energy [kcal/mol] associated with each secondary structure.

Figure 4. Schematic model of the putative recombination mechanism originating the deletion. A. Wide arrow indicates the*DMD* gene oriented towards the telomere (Xp tel, little circle), opposite to the centromere (X cen, line arrowhead). **B.** Paternal and Maternal Homologous X-chromosomes. "Black rectangles": maternal*DMD* exons; "White rectangles": paternal *DMD* exons; STR"n": where number indicate their intronic location (introns 62, 49, 45 and 44); "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 STR62 and STR44, observed from segregation analysis. **D:** Resulting III4´s chimeric X-chromosome structure, showing the deletion as the absence of STR45 and STR49.

Figure 5. Schematic representation of possible molecular mechanisms originating the SVs. A. Duplication FoSTeS mechanism. From top downwards. The 1st collapse of the replication fork may stop the DNA synthesis on the leading strand allowing the invasion of the lagging strand (on top), acting as a template to synthetize 7 bases representing an inverted insertion aided by the "biting" of the 3' nucleotide "T". A 2nd collapse stops the replication, permitting a 2nd template switching to the leading strand, near the 1st collapse´s sequence and not aided by microhomologies. This may proceed with the synthesis of an additional tract of 11 bases. Finally, a 3rd collapse may stop the synthesis at a dinucleotide "AG" on the

12

3' end, which led to a 3rd template switching, invading the lagging strand several kb upstream defining the starting point of the duplicated region. "Black star": Collapse sites of the replication fork. **B. Deletion NHEJ model.** DNA breaks located in both maternal homologous X-chromosomes were repaired by a classical route of NHEJ, where the gap structure may have been stabilized by a single bp "T" microhomology. "Black triangles": 5' and 3' DNA breakpoints of the large deletion.

Supplementary Figure 1. Schematic representation of the chimeric reads involved in the SVs. Each figure shows the key WGS chimeric reads involving the deletion (**A**) and the duplication (**B**) breakpoints. Each read is shown with their respective ID number. The reads below represent sequences target of the deletion (A) and the duplication (B) breakpoints, relevant X-chromosome coordinates (NG_012232.1, GRCh38.p13) are depicted.

A.

B.

C.

Deletion

Duplication

A.

III4 Deletion

bkp 5'   bkp 3'

ΔG=-1.95   ΔG=-3.77

B.

III4 Duplication

| 1st bkp 5' | 1st bkp 3' | 2nd bkp 5' | 2nd bkp 3' | 3rd bkp 5' | 3rd bkp 3' |

ΔG=-0.27   ΔG=-0.75   ΔG=-1.31   ΔG=-0.79   ΔG=-0.27   ΔG=-5.01

A.

ID6950

ID23109

ID9251

ID26052

ID6762

GAATTCTATTGTGAAGACTAAAGTGACTATTAAAGTAAGAAATGTTAGTAGTCAATATTCCAACCAGCTAGGGCATGAGGTCTTGGCAGA

NG_012232.1 1251463   1251505   1697983   1698029

B.

ID22075

ID18803

ID15702

ID17613

ID10128

ID30545

GATCTTTTATATTTTTGTTTCAATTTCATTTAGTTCT**TAAAATGCAATTTCATTT**AGCTCTTAAAATGAAAAGGTAAGTAAAAAAATTAAAA

NG_012232.1 1122141   1122177   990894   990929

16