



Published in final edited form as:

*Hum Mutat.* 2022 April ; 43(4): 511–528. doi:10.1002/humu.24343.

## Intron Mutations and Early Transcription Termination in Duchenne and Becker muscular dystrophy

Megan A. Waldrop<sup>1,2,3</sup>, Steven A. Moore<sup>4</sup>, Katherine D. Mathews<sup>5</sup>, Benjamin W. Darbro<sup>5</sup>, Livja Medne<sup>6</sup>, Richard Finkel<sup>7,\*</sup>, Anne M. Connolly<sup>8,\*\*</sup>, Thomas O. Crawford<sup>9</sup>, Daniel Drachman<sup>9</sup>, Nicolas Wein<sup>1</sup>, Ali A. Habib<sup>10,\*\*\*</sup>, Monika A Krzesniak-Swinarska<sup>11</sup>, Craig M. Zaidman<sup>8</sup>, James J. Collins<sup>12</sup>, Manu Jokela<sup>13,14</sup>, Bjarne Udd<sup>13</sup>, John W. Day<sup>15,\*\*\*\*</sup>, Gloria Ortiz-Guerrero<sup>16</sup>, Jeff Statland<sup>16</sup>, Russell J. Butterfield<sup>17</sup>, Diane M. Dunn<sup>18</sup>, Robert B. Weiss<sup>17,18</sup>, Kevin M. Flanigan<sup>1,2,3</sup>

<sup>1</sup>The Center for Gene Therapy, Nationwide Children's Hospital, Columbus, OH 43205

<sup>2</sup>Department of Neurology, The Ohio State University, Columbus, OH 43205

<sup>3</sup>Department of Pediatrics, The Ohio State University, Columbus, OH 43205

<sup>4</sup>Department of Pathology, The University of Iowa, Iowa City, IA, 52242

<sup>5</sup>Department of Pediatrics, The University of Iowa, Iowa City, IA, 52242

<sup>6</sup>Children's Hospital of Philadelphia, Philadelphia, PA 19104

<sup>7</sup>Nemours Children's Hospital, Orlando, FL 32827

<sup>8</sup>Department of Neurology, Washington University, Saint Louis, MO 63110

<sup>9</sup>Johns Hopkins University, Baltimore, MD 21218

<sup>10</sup>Columbia University, New York, NY 10032

<sup>11</sup>Department of Neurology, University of New Mexico Health Sciences Center, Albuquerque, NM 87131

<sup>12</sup>Department of Pediatric Neurology, Mercy Hospitals, Springfield, MO 65804

<sup>13</sup>Neuromuscular Research Center, Tampere University Hospital and University of Tampere, Tampere, Finland

**Corresponding authors:** Kevin Flanigan, MD, Abigail Wexner Research Institute, Nationwide Children's Hospital, 700 N. Children's Drive, WA3011, Columbus, Ohio 43205, kevin.flanigan@nationwidechildrens.org, Robert B. Weiss, PhD, Department of Human Genetics, University of Utah, 15 N 2030 E, Salt Lake City, Utah 84107, bob@genetics.utah.edu.

\*Current affiliation: St. Jude Children's Research Hospital, Memphis TN 38105

\*\*Current Affiliation: The Center for Gene Therapy, Nationwide Children's Hospital, Columbus, OH 43205

\*\*\*Current affiliation: University of California, Irvine, CA 92697

\*\*\*\*Current affiliation: Department of Neurology, Stanford University Medical Center, Palo Alto, CA 94304

**Author contributions:** MAW, RBW and KMF contributed to the conception and design of study, acquisition and analysis of data and drafting a significant portion of the manuscript or figures. SAM and BWD contributed to acquisition and analysis of data and drafting a significant portion of the manuscript or figures. KDM, LM, RF, AMC, TOC, DD, NW, AH, MAK-S, CMZ, JJC, MJ, BU, JWD, GO-G, JS, RJB and DMD contributed to the acquisition and analysis of data. All authors critically reviewed and revised the manuscript.

**Conflicts of Interest:** Nothing to report.

RNA-Seq data are available under the NCBI GEO Series accession number GSE175861.

<sup>14</sup>Division of Clinical Neurosciences, Turku University Hospital and University of Turku, Turku, Finland

<sup>15</sup>Department of Neurology, University of Minnesota Medical Center, Minneapolis, MN 55454

<sup>16</sup>Department of Neurology, University of Kansas, Kansas City, KS

<sup>17</sup>Department of Pediatrics, The University of Utah School of Medicine, Salt Lake City, UT 84112

<sup>18</sup>Department of Human Genetics, The University of Utah School of Medicine, Salt Lake City, UT 84112

## Abstract

*DMD* pathogenic variants for Duchenne and Becker muscular dystrophy are detectable with high sensitivity by standard clinical exome analyses of genomic DNA. However, up to 7% of *DMD* mutations are deep intronic and analysis of muscle-derived RNA is an important diagnostic step for patients who have negative genomic testing but abnormal dystrophin expression in muscle. In this study, muscle biopsies were evaluated from 19 patients with clinical features of a dystrophinopathy, but negative clinical *DMD* mutation analysis. Reverse transcription PCR (RT-PCR) or high-throughput RNA sequencing (RNA-Seq) methods identified 19 mutations with one of three pathogenic pseudoexon types: deep intronic point mutations, deletions or insertions, and translocations. In association with point mutations creating intronic splice acceptor sites, we observed the first examples of *DMD* pseudo 3'-terminal exon mutations causing high efficiency transcription termination within introns. This connection between splicing and premature transcription termination is reminiscent of U1 snRNP-mediated telescripting in sustaining RNA polymerase II elongation across large genes, such as *DMD*. We propose a novel classification of three distinct types of mutations identifiable by muscle RNA analysis, each of which differ in potential treatment approaches. Recognition and appropriate characterization may lead to therapies directed toward full-length dystrophin expression for some patients.

## Keywords

Becker muscular dystrophy; Duchenne muscular dystrophy; pseudoexon; deep intronic; transcription termination; telescripting

## 1 | INTRODUCTION

X-linked dystrophinopathies are the most common muscular dystrophies with an incidence ranging from 1:3800 to 1:6200 in newborn boys of European descent (Mendell et al., 2012), with a prevalence in 2010 estimated at 1.38 per 10 000 males aged 5 to 24 years (Romitti et al., 2015). The dystrophinopathies include Duchenne muscular dystrophy (DMD), intermediate muscular dystrophy (IMD), Becker muscular dystrophy (BMD) and the extremely rare X-linked dilated cardiomyopathy; all are caused by mutations in the *DMD* gene which encodes dystrophin. DMD is the most severe phenotype; these boys typically express no dystrophin and in the pre-corticosteroid era would lose ambulation by age 12 and die of cardiac or respiratory failure in the second or third decade of life. The designation intermediate muscular dystrophy has been used to describe boys who lose

ambulation between the ages of 12 and 15 years(Flanigan et al., 2009). Becker muscular dystrophy is associated with mutations that allow expression of a partially functional dystrophin protein and a milder phenotype. The mutational heterogeneity of BMD is associated with broad clinical variability, with loss of ambulation ranging from just beyond age 15 years to late adulthood(Flanigan, 2014).

The *DMD* gene is nearly 2.2 megabases in length with 79 exons, although 99% of the gene is intronic sequence(Muntoni, Torelli, & Ferlini, 2003). According to the reading-frame rule(Monaco, Bertelson, Liechti Gallati, Moser, & Kunkel, 1988), DMD is caused by mutations that disrupt or truncate the reading frame and result in no functional dystrophin, while BMD is caused by mutations that maintain an open reading frame and allow some functional or partially functional dystrophin to be made. However this rule accurately predicts phenotype ~90% of the time in DMD and less frequently in BMD(Aartsma-Rus, Van Deutekom, Fokkema, Van Ommen, & Den Dunnen, 2006; Flanigan et al., 2009), and its specificity may vary depending on the mutation type. For example, at one clinical center reporting on nearly 450 dystrophinopathy cases the reading frame rule was upheld for 93% of exon deletion but only 66% of exon duplications(Takeshima et al., 2010). Deletions of one or more exons account for around 65% of dystrophinopathy mutations, and exon duplications around 6–11%; subexonic mutations account for most of the remainder, and include nonsense mutations, insertions or deletions (indels), splice site mutations, and very rarely missense mutations (Aartsma-Rus et al., 2006; Beroud et al., 2005; Bladen et al., 2015; Dent et al., 2005; Flanigan et al., 2009; Takeshima et al., 2010; Tuffery-Giraud et al., 2009).

An underrecognized class of mutations is that which occurs in deep intronic non-coding regions of the gene and leads to inclusion of intronic sequence as a pseudoexon(Gurvich et al., 2008; Tuffery-Giraud, Saquet, Chambert, & Claustres, 2003; Zaum et al., 2017). Genomic DNA-based analyses used in common clinical practice fail to identify these, as most use a combination of copy number detection (via a method such as MLPA) along with sequencing of the coding region of the gene. The frequency of such intronic pseudoexon mutations is unclear, but the most relevant data for a potential range comes from surveys of unselected clinic populations, as opposed to databases aggregating reported mutations. The sensitivity of the standard clinical mutational analysis approach, utilizing DNA from blood samples, ranges between 93 and 98.5% in patients with a dystrophinopathy (with the latter value calculated from aggregate data presented in the publication of Takeshima *et al*)(Dent et al., 2005; Takeshima et al., 2010; Yan, Feng, & Buzin, 2004). This suggests that the remaining dystrophinopathy patients—up to 7%—will have pseudoexon mutations, which require analysis of muscle-derived mRNA (either via sequencing of RT-PCR generated cDNA or via RNA-Seq) for identification. Despite a  $\beta$ -globin pseudoexon being among the very first splicing mutations described for human disease (Treisman, Orkin, & Maniatis, 1983), these mutations have been underappreciated as a group due to their relative rarity. However, pseudoexons may be particularly prevalent in the dystrophinopathies because the extremely large genomic size of the *DMD* locus provides a bigger target for intronic variations.

In our ongoing studies of genotype/phenotype correlations in a large research cohort, and via associated clinical mutation analysis, we performed mRNA analysis in patients with a dystrophinopathy based on clinical features and abnormal dystrophin protein expression in muscle biopsy. As a result, we identified 19 causative deep intronic variants or related mutations resulting in RNA alterations of the *DMD* gene in these patients, further highlighting the genotypic complexity of the dystrophinopathies.

## 2 | MATERIALS AND METHODS

### 2.1 | Patients

Patients were ascertained from two sources: (1) patients enrolled in the United Dystrophinopathy Project (UDP), an ongoing natural history and genotype-phenotype database consortium, and (2) patients referred for specialized *DMD* gene testing at the University of Utah Genome Center. All patients had clinical features of dystrophinopathy or an X-linked family history of muscular dystrophy. All patients included in our analysis underwent routine clinical testing that interrogated all exons by copy number and sequencing but did not identify a mutation in the *DMD* gene. Such methods included multiplex ligation dependent probe amplification (MLPA) followed by single condition amplification/internal primer sequencing (SCAIP), or chip-based microarray methods. In most patients, the failure to identify *DMD* mutations led to diagnostic muscle biopsies; in some, the biopsy diagnosis of dystrophinopathy was established prior to the performance of the clinical genomic mutation test. Regardless, all patients were diagnosed with a dystrophinopathy based on altered or absent dystrophin expression by immunohistochemical, immunofluorescent, or immunoblot analysis of a muscle biopsy (see representative examples, Figure S1 and S2).

### 2.2 | Mutational analysis

Under institutional review board approved protocols, and following parental and/or patient consent, genomic blood and/or archived muscle tissue stored at  $-80^{\circ}\text{C}$  (generally derived from clinical biopsies) was obtained for analysis. A sufficient quantity of archived muscle tissue was available from all patients to perform diagnostic mRNA extraction and RNA-based analysis by either RT-PCR and subsequent cDNA sequencing analysis, or by RNA-Seq analysis. Because patients were ascertained between 2005 and 2019, methodology changed over time; earlier samples were assessed using RT-PCR, and later samples using massively parallel RNA Sequencing (RNA-Seq). Once a pathogenic variation was identified from mRNA analysis and mapped to genomic coordinates, sequencing of genomic DNA was performed to confirm the underlying mutation in the *DMD* locus. For each patient, intronic primers were designed that flanked the site of the pathogenic variant and spanned the pseudoexon (or similar mutation). Sequencing was performed from PCR products amplified from genomic DNA using standard Sanger sequencing techniques.

For either method of analysis, total RNA was extracted from approximately 40, 10  $\mu\text{m}$  thick cryosections of frozen muscle tissue, which were triturated in 400 microliters of 20 mM Tris\*Cl pH 7.4, 150 mM NaCl, 5 mM  $\text{MgCl}_2$ , 1 mM DTT, and 1% Triton X-100 followed by isolation with TRIzol LS (Life Technologies) according to the manufacturers'

recommended protocol. For earlier samples, the cDNA of the dystrophin gene was amplified by RT-PCR using a one-step RT-PCR kit (Life Technologies) and ten overlapping primer sets (Roberts et al. 1991). RT-PCR reactions were used as template in a second round of nested PCR using Expand High Fidelity Kit (Roche). PCR products were visualized on an ethidium bromide-stained gel after electrophoresis to confirm amplification before being cleaned with ExoSAP-IT® (USB) according to the manufactures' recommended protocol. All products were sequenced (Eurofins/Operon), aligned to a consensus sequence (NM\_004006.2) and analyzed using Sequencer 5.0® (Gene Codes) software. After identification of pseudoexons in the cDNA, specific primers were designed to identify the DNA level mutations for each patient. Intronic primer sequences used for amplification of genomic DNA are available upon request.

For RNA-Seq, libraries were constructed from ~1 microgram of total RNA depleted for cytoplasmic and mitochondrial rRNA using the tiling oligodeoxynucleotides/RNase H digestion method (Adiconis et al., 2013). Random-primed, indexed RNA-Seq libraries were prepared using Illumina TruSeq Stranded Total RNA library kits and sequenced on an Illumina HiSeq 2500 instrument using either single-end 50 bp or paired-end 125 bp v4 read chemistry. BAM alignment files were generated by mapping quality trimmed FASTQ sequence reads to the GRCh37/hg19 human (Feb. 2009) reference genome using the STAR v2.7 RNA-Seq aligner (Dobin et al., 2013) with the mapping parameter set: `--alignSJDBoverhangMin 1 --alignSJoverhangMin 8 --outFilterMismatchNoverLmax 0.3 --chimSegmentMin 15 --outFilterMultimapNmax 10 --alignIntronMax 1100000 --twopassMode Basic`.

Splicing analysis of *DMD* transcripts was performed by visualizing splice junction reads in the IGV browser using strand-specific coverage plots and sashimi plots generated from BAM alignment files. Additional coverage analysis used the BEDTools genomecov function with the UCSC bedGraphToBigWig utility and annotated sashimi plots of the splicing events were generated with the ggsashimi tool (Garrido-Martin, Palumbo, Guigo, & Breschi, 2018). Read counts spanning novel junctions were refined by STAR re-mapping using the `-sjdbFileChrStartEnd` parameter with a list of novel splice junction coordinates.

### 2.3 | *DMD* intron coverage analysis

Read counts for *DMD* intron coverage analysis were summarized using Dp427m transcript annotations either as individual introns or as non-overlapping 5-kilobase intronic windows from chrX:31140097–33229348. These gene and feature annotations were used to extract read count summaries from the STAR-aligned BAM files using the featureCounts tool (Liao, Smyth, & Shi, 2014). For local normalization, read counts from each 5 kb intronic segment were normalized to total *DMD* intronic read counts and a  $\log_2$  fold change was calculated by comparing each segment to locally normalized *DMD* read coverage from control muscle biopsies. Alternatively, for global transcriptome normalization, the read count summaries for individual *DMD* exons and introns were combined with transcriptome-wide gene-level summaries. For this analysis, the 125 nt. paired end libraries were trimmed to 50 nt. single-end, stranded libraries to make read mapping comparable between samples. After STAR alignment, read counts were extracted using featureCounts for protein coding genes

from the GENCODE project lifted annotations from V31lift37 (Ensembl 97). Exon and intron-level read count summaries for *DMD* were also extracted with featureCounts and combined with gene-level summaries. Read counts were normalized using the TMM method from the *edgeR* Bioconductor package without filtering to remove low count genes. Library size per sample averaged 19.3 million read counts (range 11.4 to 30.0 million) for muscle biopsy RNA-Seq samples and the *edgeR* cpm function was used to compute normalized counts per million (cpm) or  $\log_2$  cpm values used for fold change analysis. Patient versus control fold change ratios for individual *DMD* exons/introns were calculated using the  $\log_2$  cpm values, and only introns > 2.5 kb in length were analyzed to ensure sufficient read counts. The control RNA-Seq libraries were from an unaffected 9-year-old male for the pairwise comparison; and from 2, 8, 27, and 28-year-old males.

## 2.4 | Splice site strength analysis

Splice site strength predictions utilized the following tools: MaxEntScan (<http://hollywood.mit.edu/burgelab/maxent/Xmaxentseq.html>), Human Splicing Finder (HSF3.1, <http://www.umd.be/HSF3/HSF.shtml>), ESEfinder3.0 (<http://krainer01.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home>), and the SpliceAI deep learning algorithm (Jaganathan et al., 2019). Bioinformatic prediction of intronic polyadenylation sites utilized the IPAFinder\_PS\_FET.py program from IPAFinder and the POLYAR program with parameters for PAS-strong poly(A) sites (Akhtar, Bukhari, Fazal, Qamar, & Shahmuradov, 2010; Zhao et al., 2021). Genome-wide, pre-computed SpliceAI scores were downloaded from <https://basespace.illumina.com/s/5u6ThOblecrh> and variants from chrX:31140097–33229348 with splice acceptor or donor scores greater than 0.1 were extracted. The SpliceAI predicted scores for *DMD* region variants and individual mutations were confirmed by re-analysis with SpliceAI run as a local instance using software available from <https://github.com/Illumina/SpliceAI>. The ggseqlogo R package was used to display sequence alignments of splice acceptor and donor sites using the probability method for height scaling.

## 3 | RESULTS

### 3.1 | Classification of *DMD* intronic mutations

A total of 1,344 subjects from the UDP and 1,259 clinical tests had positive *DMD* mutations. Out of those samples, we identified 19 patients with clinical features, dystrophic muscle biopsies (Supp. Figures 1 and 2), and muscle dystrophin expression consistent with a dystrophinopathy who had deep intronic or other pathogenic variants resulting in altered *DMD* RNA transcripts. In each case, RNA analysis was performed after standard clinical *DMD* gene testing was negative. For the purposes of characterization as outlined in Table 1, we describe three general types of mutations: Type 1, consisting of point mutations; Type 2, consisting of multi-nucleotide deletions or insertions; and Type 3, consisting of larger chromosomal rearrangements/translocations.

As shown in Table 2 and Figure 1, the largest group consisted of point mutations (Type 1), a type of pseudoexon-creating mutation that has been previously identified and well-characterized (Baskin, Gibson, & Ray, 2011; Bovolenta et al., 2008; Cagliani et al., 2004;

Ginsberg, McCarty, Lacomis, & Abdel-Hamid, 2018; Gonorazky et al., 2016; Greer et al., 2015; Gurvich et al., 2008; Jones et al., 2019; Khelifi et al., 2011; Madden, Fletcher, Davis, & Wilton, 2009; Magri et al., 2011; Oshima et al., 2009; Zaum et al., 2017). In this cohort, we found eight Type 1a point mutations that created splice donor or acceptor sites (Figure 1A–H), two Type 1a point mutations that removed a decoy splice acceptor site (Figure 1I,J), one Type 1b that creation/disruption an exon splice enhancer/silencer motif (Figure 1K) leading to the utilization of cryptic splice donor and acceptor signals, and a novel class of two Type 1c point mutations that created splice acceptor sites (Figure 1L and M) leading to the formation of pseudo 3'-terminal exons. As expected, these pseudoexon-activating point mutations were embedded in intronic sequences that resembled splice acceptor and donor consensus, as seen by comparing the Type 1 pseudoexon splice site sequence alignments to constitutive splice sites from the *DMD*Dp427m transcript isoform (Figure 1N).

Three patients had Type 2 mutations (Figure 2A–C) consisting of multi-nucleotide deletions or insertions. One patient with DMD had a 34 nt deletion within intron 7 (c.650–916\_650–883del) that resulted in flanking sequences being included as a novel out-of-frame pseudoexon that utilized an upstream acceptor and downstream donor site. We analyzed this pseudoexon using the HSF3.1 program, which identified that the novel junction sequence may have resulted in enhanced utilization of cryptic splice donor and acceptor sites by two complementary mechanisms. First, it creates two new potential exon splice enhancer sites, recognized by two different SR proteins important to splicing: one site is recognized by the SRSF2 (formerly SC35) protein (UGCUACUA), and one is recognized by the SRSF5 (formerly SRp40) protein (CUACUAG); in addition, it is predicted to disrupt an exon splice suppressor site (UCAGAACU). Also, we analyzed this mutation using SpliceAI algorithm which predicts delta ( ) acceptor and donor scores ranging from 0 to 1 between a reference and alternate allele and can be interpreted as the probability of the variant being splice-altering. The SpliceAI-10k algorithm predicted that this 34 nt deletion activates strong splice acceptor ( score = 0.61) and donor ( score = 0.76) sites at the observed locations (Figure 2A). A second Type 2 mutation (Figure 2B) found in a BMD patient, had a 238 nt deletion that spanned the exon 45/intron 45 junction (c.6489\_6614+111del); the resulting hybrid exon, which maintains an open reading frame, contains 48 nt from the first part of exon 45 and 30 nt from intron 45 beginning at a position +110 relative to the end of exon 45. A third Type 2 mutation (Figure 2C) found in a patient with DMD, showed a L1 retrotransposon element (LINE) insertion entirely within exon 33 (c.4606\_4607ins10218+As). The resulting hybrid pseudoexon contained the first 88 nt from exon 33, and 95 nt from the 5' portion of the LINE element insertion, spliced to exon 34, and results in translational termination 17 amino acids downstream in the new reading frame.

Patients with Type 3 mutations represented larger scale chromosomal rearrangements, and all had a DMD phenotype. One patient had an annotated chr. 2 lncRNA exon (exon 2 from NR\_131227.1) detected by RT-PCR analysis (Figure 2D), indicating a portion of that chromosome 2 genomic region had been inserted into *DMD* intron 55. RNA-Seq analysis resolved the final two cases as a transposition of the chromosome X *OFD1* region into *DMD* intron 1 (Figure 2E) and an inversion of the short arm of chromosome X (Figure 2F). For one patient, RNA-Seq detected normal levels of intron 1 *DMD* reads that ended abruptly near chrX:33,112,600 (Figure 2G). An intron *DMD* g.33112674 :: *OFD1* g.13782326

breakpoint was detected both by RNA-Seq reads and by PCR from genomic DNA both in the proband and his mother (Figure 2G). A chromosomal inversion was not observed by cytogenetic staining of chromosomal spreads, but cytogenetic microarray analysis detected duplications (Supp. Figure 3) both in *DMD* intron 1 (chrX:~33100000–33150000) and in the *OFDI* region (chrX:~13720000–13780000), suggesting a complex transposition event. Transcription from the *DMD* intron 1 into the transposed region proceeded in the antisense orientation into the *OFDI* gene and eliminated most transcription through the *DMD* gene, probably due to the insertion of a strong transcriptional terminator from the *OFDI* region (Figure 2H). RNA-Seq analysis of the final Type 3 patient indicated an ~23 Mb inversion breakpoint, confirmed by PCR from genomic DNA, between *DMD* intron 60 and the *VCX2-VCX3B* region at chrX:8200000, where transcription from the *DMD* region proceeded for an additional ~300 kb (Figure 2I).

A summary of patient clinical and genotypic features is provided in Table 2. Of the 19 patients reported, 10 had the milder BMD phenotype. In this group, symptom onset occurred as early as 1 year and as late as 56 years of age with all ambulant at the time of this report (age 7–68 years). All but one BMD patients had normal cardiac function, and varying degrees of preservation of normal splicing were observed in 4 out of these 10 patients analyzed by RNA-Seq (Table 2, Figure 3). All of the DMD patients had typical features, with symptom onset ranging from 6 months to 3 years and loss of ambulation between 9 and 12 years. Only one DMD patient had cardiomyopathy, although this is a young cohort and all of those with reported normal cardiac function were age 13 or younger. Consistent with the more severe phenotype, 8 of the 9 DMD patient samples showed only out-of-frame RNA transcripts, with no significant wild-type transcript. The sole exception is the subject whose pseudoexon consisted of a 51 nt fragment of chromosome 2 sequence, in whom the resulting translated protein might be expected to be significantly misfolded and unstable (p.Glu2739\_Asp2740ins17).

### 3.2 | Activation of intronic splice acceptor sites cause premature transcription termination

Splice junction reads from two BMD patients with Type 1c point mutations each activated an intronic splice acceptor site that led to inclusion of a pseudo 3'-terminal exon and unexpectedly caused premature termination of the transcribing RNA polymerase complex. Unlike other pseudoexons detected by RNA-Seq analysis (Supp. Figure 3), downstream splice donor site junction reads were not observed for the intron 43 c.6290+3076A>G (Figure 3A) or the intron 61 c.9163+2510G>A (Figure 3B) mutations, and elevated intronic read depths persisted for at least 5 kb downstream of these mutations. To identify the locations of putative intronic polyadenylation sites at the 3' border of these pseudo-terminal exons, we applied the IPAFinder algorithm which performs *de novo* identification of intronic poly(A) events using splice junction reads and intronic RNA-seq read coverage. IPAFinder predicted a terminal exon (located at chrX:31364161–31358615) from a paired analysis of the intron 61 g.31364162, c.9163+2510G>A mutation versus a normal control (*P* value from Fisher's exact test = 7e-05). The 3' end of predicted 5.5 kb intron 61 pseudo-terminal exon coinciding with a precipitous decline in read depth; independently, the POLYAR program predicted a strong polyadenylation site at this putative 3' border of this pseudo-terminal



exon (Figure 3C). In contrast to the intron 61 mutation, IPAFinder did not predict the location of an intronic poly(A) site for the intron 43 mutation, suggesting that the more gradual decline of intron 43 read coverage may reflect the combined use of multiple sites. To test if these activated intronic splice acceptor sites were coupled to premature transcription termination, we used two different approaches for measuring normalized levels of RNA-Seq reads from introns downstream of these mutations. First, read counts were summarized in non-overlapping 5 kb intronic intervals across the entire ~2.1 Mb span of the Dp427m isoform and interval coverage depths were calculated relative to total *DMD* intronic read counts mapped in that patient. The ratio of these coverage values in the patient versus an unaffected 9-year-old male were plotted across the *DMD* locus as 'locally' normalized  $\log_2$  fold change values (Figure 4A). For both Type 1c mutations, the read depth within the intron containing the mutation was increased followed by an abrupt decrease in intron coverage 3' of the mutation; this pattern of decreased intronic read depth was not apparent with other pseudoexon point mutations (Supp. Figure 4A). The decreased intronic read depth pattern downstream of the mutation is reminiscent of RNA-Seq coverage in genes that undergo telescripting, where failure of U1 snRNP to suppress premature RNA polymerase II termination leads to abrupt decreases in intronic coverage after sites of premature termination (So et al., 2019).

RNA polymerase II prematurely terminating after transcribing across these Type 1c point mutations would also decrease the relative abundance of exons downstream of the mutation. To test this second prediction, *DMD* exon and intron read counts were normalized to the entire protein coding transcriptome and relative fold changes were calculated as the ratio of counts per million values from the patient versus a 9-year-old, non-dystrophic male. These 'globally' normalized *DMD* exon and intron mRNA fold change values are plotted in their 5' to 3' direction of transcription in Figure 4B. The normalized *DMD* intron pre-mRNA levels for the c.6290+3076A>G (intron 43) mutation were unchanged from intron 1 through 42, followed by a slight elevation in intron 43 level and then a sharp reduction (~5-fold) from intron 44 through 62, then elevating to intermediate levels from intron 63 through 78 coinciding with low-level transcription from the Dp71 promoter in intron 62 (Figure 4B, right panel). The normalized *DMD* exon pre- and mature mRNA levels for this mutation closely paralleled the intron levels with a sharp reduction in mRNA levels after exon 43 to exon 62 and then slightly elevated with the Dp71 transcript beginning at exon 63. This parallel reduction in exon and intron mRNA levels in the vicinity of the intron 43 mutation strongly supports the model for premature termination of RNA polymerase II transcription in the vicinity of the activated intronic splice acceptor site. The *DMD* exon and intron mRNA levels for the c.9163+2510G>A (intron 61) mutation were unchanged from exon 1 through 60, followed by a ~5-fold increase in intron 61 and a ~5-fold decrease in intron 62 coverage consistent with pseudo 3'-terminal exon formation within intron 61; intermediate levels beginning at exon 63 through 79 indicated Dp71 expression in this patient's muscle biopsy, as well (Figure 4B, right panel). In contrast, globally normalized *DMD* intron mRNA levels from other pseudoexon point mutations (Supp. Figure 4B) showed only minor fluctuations across all 78 introns, as did RNA-seq data from additional non-dystrophic controls ranging in age from 2 to 28 years old (Supp. Figure 4C). At the exon mRNA level, these other pseudoexon point mutations showed a relative 5' to 3' decrease in *DMD*

exon mRNA, particularly for the intron 7 c.650–39498A>G, intron 38 c.5448+67A>G, and intron 47 c.6913–5879A>G mutations (Supp. Figure 4B). This progressive 5' to 3' decrease in *DMD* exon mRNA levels has been previously observed in *DMD* patients using RT-qPCR and was interpreted as post-transcriptional nonsense-mediated decay (NMD) of the mature mRNA transcript overlaid with pre-mRNA transcript accumulation differences in the 5' - versus 3' -ends of the gene due to the additional nuclear half-life of 5' exons during the 16 hours required to transcribe the ~2.1 Mb Dp427m pre-mRNA (Anthony et al., 2014; Tennyson, Shi, & Worton, 1996). A similar 5' to 3' exon mRNA decrease has been observed in the *mdx* mouse (exon 23 nonsense mutation) and in myotubes derived from a patient with an out-of-frame exon 48 to 50 deletion; that study suggested that a decrease in *DMD* transcription instead of cytoplasmic NMD was responsible for reduction in *DMD* mRNA (Garcia-Rodriguez et al., 2020). For the pseudoexon mutations shown in Figure S4, the lack of a parallel 5' to 3' decrease in intron mRNA levels suggests that the 5' to 3' decrease in exon mRNA levels is due to a post-transcriptional mechanism, most likely NMD. However, the corresponding decrease of exon and intron mRNA read depths downstream of the Type 1c intronic splice acceptor site mutations support the hypothesis that splicing to these mutation-activated acceptor sites results in 3'-end cleavage of the pre-mRNA and premature transcription termination within the *DMD* gene in close proximity to the mutation.

### 3.3 | Intronic target size estimation for pseudoexon activation

We compared the 13 pseudoexon point mutations found in this study to 24 *DMD* pseudoexon mutations previously described in the literature and in the ClinVar database (Supp. Table 2) and found only three mutations in common: c.650–39498A>G, c.3603+820G>T, and c.9225–647A>G. Since the SpliceAI algorithm accurately predicted the observed pseudoexon location for the Type 2 c.650–916\_c.650–883del mutation (Figure 2A), we calculated SpliceAI scores for the Type 1a, 1b and 1c point mutations, as well as the previously described *DMD* pseudoexon mutations. The locations of the pseudoexon boundaries were accurately predicted for each Type 1a and 1b mutation, and the mean SpliceAI score was 0.66 for acceptor gain and 0.71 for donor gain for Type 1a mutations, where a score > 0.5 is considered a high confidence prediction (Jaganathan et al., 2019), while the Type 1c pseudo 3'-terminal exons showed only high scores for acceptor gain (Supp. Table 1). The pseudoexon locations of the previously known mutations (23 Type 1a and one Type 1b) were also accurately predicted, with mean scores of 0.65 for acceptor gain and 0.62 for donor gain (Table S2). The pseudoexon location of Type 1a intron 22 c.2949+909C>T (exon 22a +21C>T) mutation was correctly predicted although without high confidence scores for acceptor and donor gain (Table S1). While the SpliceAI deep learning algorithm does not specifically report what sequence features contributed to the accurate definition of exon 22a, RNA binding protein motif analysis surrounding the +21 C>T mutation suggested both loss of MBNL1/HNRPNK/SRSF1 motifs and gain of an ETR-3 motif may have been contributing factors. Since the SpliceAI predictions displayed good sensitivity for the observed pseudoexons, we used these predictions as a surrogate for estimating the 'target size' for pseudoexon mutations by examining the acceptor and donor gain scores at all possible single nucleotide variants (SNVs) for the ~2.1 million *DMD* intronic nucleotide sites, excluding  $\pm 40$  nt flanking each exon. Only 13,129 variants

(0.21% of all possible variants) had relatively permissive SpliceAI acceptor or donor gain scores  $> 0.1$ , and the score distribution for these sites is shown in Figure 5A, including scores for observed pseudoexon point mutations (Jaganathan et al., 2019). As expected, the positions of these variants relative to the predicted pseudoexons were enriched in the flanking acceptor and donor regions (61.5%, Supp. Figure 5). Using the lower quartile score of the observed mutations as a threshold (0.33 for acceptor gain and 0.38 for donor gain), only the top 1,090 predicted sites ( $\sim 0.02\%$  of all possible variants) occurred in this scoring range and were highly enriched in  $-20$  to  $+2$  acceptor consensus region (27.1%) and the  $-2$  to  $+6$  donor consensus region (62.7%) flanking the predicted pseudoexons. This suggests that within *DMD* introns spanning 2.1 Mb, the effective pseudoexon target size for high penetrance point mutations is  $\sim 1$  kb, although insertion/deletion mutations will also contribute to additional targets for pseudoexon activation.

The SpliceAI predicted sites were seen throughout introns but their fine-scale distribution suggests local enrichment around proto-pseudoexons. Two extreme examples of SpliceAI site density are shown in Figure 5B and C, where in the first case, pseudoexon 7i has only one predicted SpliceAI site in the local vicinity ( $\pm 1.5$  kb) and the pseudoexon mutation (c.650–39498A>G) occurred at this site. In contrast, pseudoexon 22i (Figure 1K) has 7 predicted SpliceAI sites in the local vicinity ( $\sim 300$  bp), one of which coincides with the observed ESE mutation (c.2949+909C>T), while 6 additional SpliceAI sites would generate the equivalent pseudoexon and 8 sites would generate a nested set of pseudoexons sharing the same acceptor site. Previous work using ultra-deep, targeted RNA-Seq analysis of *DMD* transcripts in immortalized human muscle cell lines detected low-level splicing to intronic sites with adjacent 3' and 5' splice site motifs suggestive of 'zero-length exons', also known as recursive splicing sites, and predicted 145 recursive splicing (RS) sites within *DMD* introns (Gazzoli et al., 2016; Sibley et al., 2015). Although recursive splicing is seldom seen in human introns, an additional analysis using targeted *DMD* RNA-Seq from human skeletal muscle biopsies found overlap between these proposed RS sites and the splice junctions of three low-level alternative cassette exons (Bouge et al., 2017). A recent review of *DMD* pseudoexons proposed a more extensive concordance of 145 predicted RS sites and pseudoexon splice sites (Keegan, 2020). We tested the degree of concordance by intersecting the locations of these 145 predicted 5'RS and 3'RS sites with the 68 pseudoexon splice junctions from the new and known pseudoexons described in this study (Supp. Tables 1 and 2), and found overlap between three pseudoexon splice donor sites and three 3'RS sites and one pseudoexon splice acceptor site and a 5'RS site (Supp. Table 3). Intersection of the 145 RS sites with the top 1,090 or the complete 13,129 set of SpliceAI predicted splice acceptor or donor sites with high scores (Figure 5A) resulted in 28 (19%) and 74 (51%) RS sites overlapping this class of SpliceAI sites, suggesting that the RS sites are derived from the same intronic territory as pseudoexons.

## 4 | DISCUSSION

The current gold standard of clinical mutation analysis—exon deletion/duplication analysis with reflex to exon sequencing as needed—will successfully identify the causative mutation in most, but not all, dystrophinopathy patients. The actual frequency of such mutations is not known but can be estimated. Notably, many large collections of *DMD* mutations

are obtained from databases of established mutations, not cohort studies(Aartsma-Rus et al., 2006; Bladen et al., 2015). Among the 2603 total patient samples in our cohorts with identified *DMD* mutations (1344 from the UDP and 1259 from the clinical testing program at Utah Genome Center) with identified *DMD* mutations, 19 (0.73%) had pseudoexon mutations, a value we consider to be the lower boundary of possible frequency. A more meaningful estimate of frequency may be provided by the observation that pseudoexon mutations were found in all 19, or 100%, of all patients for whom (i) a dystrophinopathy was detected by protein expression, and (ii) sufficient tissue was available for the RNA analyses we describe. Because unbiased cohort-based studies of dystrophinopathy patients as defined by either X-linked family history or dystrophin expression abnormalities show that 1.5% to 7% of *DMD* mutations are not detectable by genomic DNA analysis (Dent et al., 2005; Takeshima et al., 2010; Yan et al., 2004), we can infer an upper boundary of an estimated frequency as 7% of all mutations. We anticipate that these estimates will be refined by improved recognition of this mutation class, and increased availability of clinical RNA sequencing to facilitate its recognition.

In an effort to better characterize these unusual mutations, we propose a classification system for *DMD* mutations identified after muscle mRNA analysis (Table 1). This classification is beneficial because it highlights and simplifies these unusual mutational mechanisms. This will also allow for ease in identification of patients with pseudoexons amenable to current therapeutic strategies, ultimately impacting diagnosis and counseling for *DMD* and *BMD* patients. Type 1 pseudoexon mutations comprise the most common previously reported type: point mutations leading to creation of splice donor, acceptor, or enhancer sites(Baskin et al., 2011; Bovolenta et al., 2008; Cagliani et al., 2004; Ginsberg et al., 2018; Gonorazky et al., 2016; Greer et al., 2015; Gurvich et al., 2008; Khelifi et al., 2011; Madden et al., 2009; Magri et al., 2011; Oshima et al., 2009; Zaum et al., 2017). Although exon-skipping therapies are perhaps most widely associated with skipping of native exons, it has also been successful in skipping pseudoexons via antisense(Bolduc et al., 2019; Gurvich et al., 2008; Rendu et al., 2013) or gene editing approaches. Type 1 mutations include those potentially amenable to splice site modification that could result in exclusion of the pseudoexon from the mature mRNA.

All mutations reported here are novel, except for three Type 1a mutations. The c.650–39498A>G mutation was previously reported as creating a cryptic splice donor site within intron 7(Zaum et al., 2017). In that case, the authors posited that the minor G allele of a known intronic polymorphism (rs113593006 G/T) was important in the development of a cryptic splice acceptor site and activation of this pseudoexon(Zaum et al., 2017). However, the patient reported here has the rs113593006 T allele and also has activation of the pseudoexon; therefore, the minor rs113593006 allele is not required for the development of the pseudoexon. Two cases of acceptor site variants, one novel and one known, both altered an AG dinucleotide at the –19 position, which improved the predicted acceptor site strength (Supp. Table 1) and created an AG-exclusion zone between the pseudoexon 3' ss AG and its upstream branch point by removing the decoy AG at position –19(Wimmer et al., 2020). An additional novel c.5448+67A>G mutation created a new donor site that redefined exon 38 splicing to the extent of causing a *DMD* phenotype. Both these types of mutations reinforce the role of exon definition as one of the earliest steps in splice site recognition.

Application of the SpliceAI deep learning algorithm to the pseudoexon mutations in this study and from the literature showed that pseudoexon activation can be accurately predicted from primary sequence variation alone. The proportion of observed mutations activating donor versus acceptor sites was similar to the distribution seen from the high confidence SpliceAI predictions, and only a small fraction of the observed mutations occurred outside of the core donor and acceptor sites. The skew in the predicted activation strength of the observed mutations also allowed us to estimate that the target size for pseudoexon point mutations is on the order of only ~1000 nucleotides within the 2.1 Mb expanse of *DMD* intronic sequence. This effective target size is consistent with the frequency of pseudoexon mutations across the spectrum of all *DMD* mutations(Flanigan et al., 2009; Tuffery-Giraud et al., 2009).

Perhaps the most unexpected observation in this study were two point mutations that activated splice acceptor sites and led to pseudo 3'-terminal exon formation coupled to premature transcription termination. Pseudoexon mutations that create 3'-terminal exons have not, to our knowledge, been previously described and were not detected in previous RNA-Seq surveys of muscle disorder transcriptomes(Cummings et al., 2017; Gonorazky et al., 2016; Waddell et al., 2021). The distinctive RNA-Seq pattern of a uniform decrease in intronic read depth downstream of the pseudo 3'-terminal exon mutations has been previously observed in two different experimental contexts. It is seen in antisense oligonucleotide (ASO) experiments that block U1 snRNA (U1) base-pairing to splice donor sites, which interferes with U1's telescripting activity in suppressing premature 3'-end cleavage and transcription termination. It is also seen in ASO experiments that target introns and direct RNase H cleavage, leading to premature transcription termination downstream of the intronic cleavage site(Lai, Damle, Ling, & Rigo, 2020). A mechanism for splicing control of transcription termination has been widely appreciated due to the normal coupling of terminal exon splicing and 3' end processing of pre-mRNA. It is well known that mutations inactivating the splice acceptor site of 3' terminal exons coordinately disrupt splicing, 3' end cleavage, polyadenylation and RNA polymerase II (Pol II) transcription termination(Dye & Proudfoot, 1999). The pseudo 3'-terminal exons observed here may be the reverse of these coupled processes where mutations that activate cryptic splice acceptor sites induce cryptic cleavage and polyadenylation (CPA) sites and subsequent premature termination of elongating RNA polymerase II (Pol II). That both of these acceptor site mutations were associated with long (>5 kb), variable pseudo 3'-terminal exons suggests an additional mechanism that may result in terminal exon formation within large, constitutively spliced introns. Recent observations suggest that U1 snRNA 5'end base-pairing to pre-mRNA splice donor sites suppresses recognition of cryptic polyadenylation signals in large introns and prevents premature transcription termination(So et al., 2019). The absence of a strong splice donor site downstream of the activated splice acceptor site may be one sequence feature contributing to the formation of pseudo 3'-terminal exons by preventing formation during the exon definition step of a stable U1 snRNP pre-mRNA complex trailing the elongating Pol II. Several genes encoding CPA and termination factors, including CSTF3 and PCF11, have an analogous mechanism using weak splice donor sites coupled to intronic polyadenylation and premature transcription termination to autoregulate their levels(Wang, Zheng, Wei, Ding, & Tian, 2019).

This splicing-dependent activation of premature transcription termination may also have an unsuspected role in pre-mRNA quality control if mis-splicing events also activate this premature Pol II termination pathway and may have implications for therapeutic strategies designed to bypass splicing mutations. The concordance of recursive splicing sites and pseudoexon splicing sites that we and others have observed may also be related to quality control of mis-splicing events. One hallmark of efficient recursive splicing is a multiple ‘sawtooth’ RNA-Seq read depth pattern within single long introns due to co-transcriptional splicing of recursive segments followed by relatively rapid degradation of the excised intronic RNA; however, *DMD* long introns display a single ‘sawtooth’ pattern characteristic of non-recursive, co-transcriptional splicing. *In vivo* skipping of a duplicated mouse *Dmd* exon 2 using AAV-U7 mediated antisense treatment converts intron 1 and intron 2 ‘sawteeth’ into a single intron 1 + 2 sawtooth, consistent with non-recursive, co-transcriptional splicing of exon 1 to exon 3 in the treated mice (Wein et al., 2014). Therefore, it is unlikely that efficient recursive splicing is used in *DMD* introns but instead the predicted RS sites may represent a salvage pathway for mis-spliced transcripts. Since recursive splicing in vertebrates acts through an ‘exon definition’ mechanism, the novel association we observed between the increased density of pseudoexon-activating SpliceAI sites at putative RS-exons suggests that the rare RS splice junction products may result from mis-spliced intermediates at intronic sites that most closely mimic exon definition rules. These mis-spliced transcripts may be resolved through either a recursive splicing salvage pathway to produce a non-aberrant mature transcript or by a salvage pathway using the novel pseudo 3’-terminal exon mechanism described here that would couple mis-splicing to premature transcription termination, discontinuing the synthesis of an aberrant transcript.

In contrast to Type 1 mutations, the more complex Type 2 and 3 mutations would not necessarily be amenable to intron sequence directed exon skipping and gene editing strategies. The Type 2 mutations described here are all unique. Although a LINE insertion deep within intron 51 resulting in a pseudoexon has been reported in one patient with BMD (Goncalves et al., 2017), and a similar intron 13 LINE insertion has been reported to result in a pseudoexon in a dog model (Smith et al., 2011), here we report the first example of a LINE insertion within an exon (exon 33) resulting in a hybrid pseudoexon (c.4606\_4607ins10218+As). The Type 3 mutations are larger scale chromosomal rearrangements which, in general, have previously been reported in dystrophinopathy (Flanigan et al., 2011), but the inversions described here provide additional insights into mechanisms that contribute to complex structural mutations. Large intragenic *DMD* single and multi-exon deletions and duplications are thought to occur through repair of double stranded breaks using non-homologous end-joining repair pathways and replication-dependent mechanisms (Ankala et al., 2012; Ishmukhametova et al., 2012; Mitsui et al., 2010). The association of the *DMD-OFD1* transposition with a local duplication in *DMD* intron 1 may reflect the use of similar DNA damage response pathways.

Among the nine patients with DMD and the ten patients with BMD/hyperCKemia, the reading frame rule applied 89% and 10% of the time, respectively. This discrepancy was investigated further in one patient with a Type 1b “out-of-frame” mutation who was essentially asymptomatic with only an elevated CK. RNA-Seq revealed the presence of two transcripts, one with the pseudoexon (55%) and one with an essentially normal transcript

(45%) (data not shown). Thus, his milder phenotype was supported by dystrophin protein production from the second transcript, and it is likely the other “out-of-frame” BMD patients have a similar mechanism. It has been previously reported that even a relatively low level of dystrophin expression can significantly ameliorate phenotype (de Feraudy et al., 2021; Kinane et al., 2018; Mendell et al., 2016; Mendell et al., 2013; Waldrop et al., 2018) and these results are further supported here.

We note that pseudoexon mutations are not limited to the dystrophinopathies, as similar mutations have been described in other forms of muscular dystrophy (e.g., the *COL6A1*, *CAPN3* and *DYSF* genes) (Blazquez et al., 2013; Bolduc et al., 2019; Dominov et al., 2014), where utilization of the pseudoexon classification system may also prove useful. In the current era of therapeutic development for the dystrophinopathies, obtaining a definitive molecular diagnosis is key for patients and their families. In addition to potential therapeutic implications, a definitive molecular diagnosis facilitates accurate genetic counseling and diagnostic testing of family members. This work highlights that in the proper clinical context, failure to identify a mutation based on genomic analysis from blood samples is insufficiently sensitive to exclude a diagnosis of dystrophinopathy, and muscle biopsy for dystrophin immunostaining and for mRNA analysis remains an essential next diagnostic step.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements:

The authors wish to acknowledge the technical assistance of L.E. Taylor and F. Gumienny, and additional assistance of J. Dalton, R. Alles, and B. Connor. The authors are grateful to Dr. T.A. Vetter for additional image preparation. This work was partially supported by NINDS grant R01 NS043264 to K.M.F. and R.B.W. S.A.M. and K.D.M. are partially supported by NINDS grant U54 NS053672 for the Iowa Wellstone Muscular Dystrophy Specialized Research Center.

## Funding Support:

This study was supported by the National Institutes of Health NINDS grants NS043264 and NS085238

## Data Availability Statement:

All identified mutations were submitted to Global Variome shared LOVD (<https://www.lovd.nl>) and are publicly available with accession numbers: 00362268, 00375246–00375251, 00375253–00375256, 00375258–00375262, 00375270–00375272.

## REFERENCES

- Aartsma-Rus A, Van Deutekom JC, Fokkema IF, Van Ommen GJ, & Den Dunnen JT (2006). Entries in the Leiden Duchenne muscular dystrophy mutation database: an overview of mutation types and paradoxical cases that confirm the reading-frame rule. *Muscle Nerve*, 34(2), 135–144. doi:10.1002/mus.20586 [PubMed: 16770791]
- Adiconis X, Borges-Rivera D, Satija R, DeLuca DS, Busby MA, Berlin AM, ... Levin JZ. (2013). Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat Methods*, 10(7), 623–629. doi:10.1038/nmeth.2483 [PubMed: 23685885]

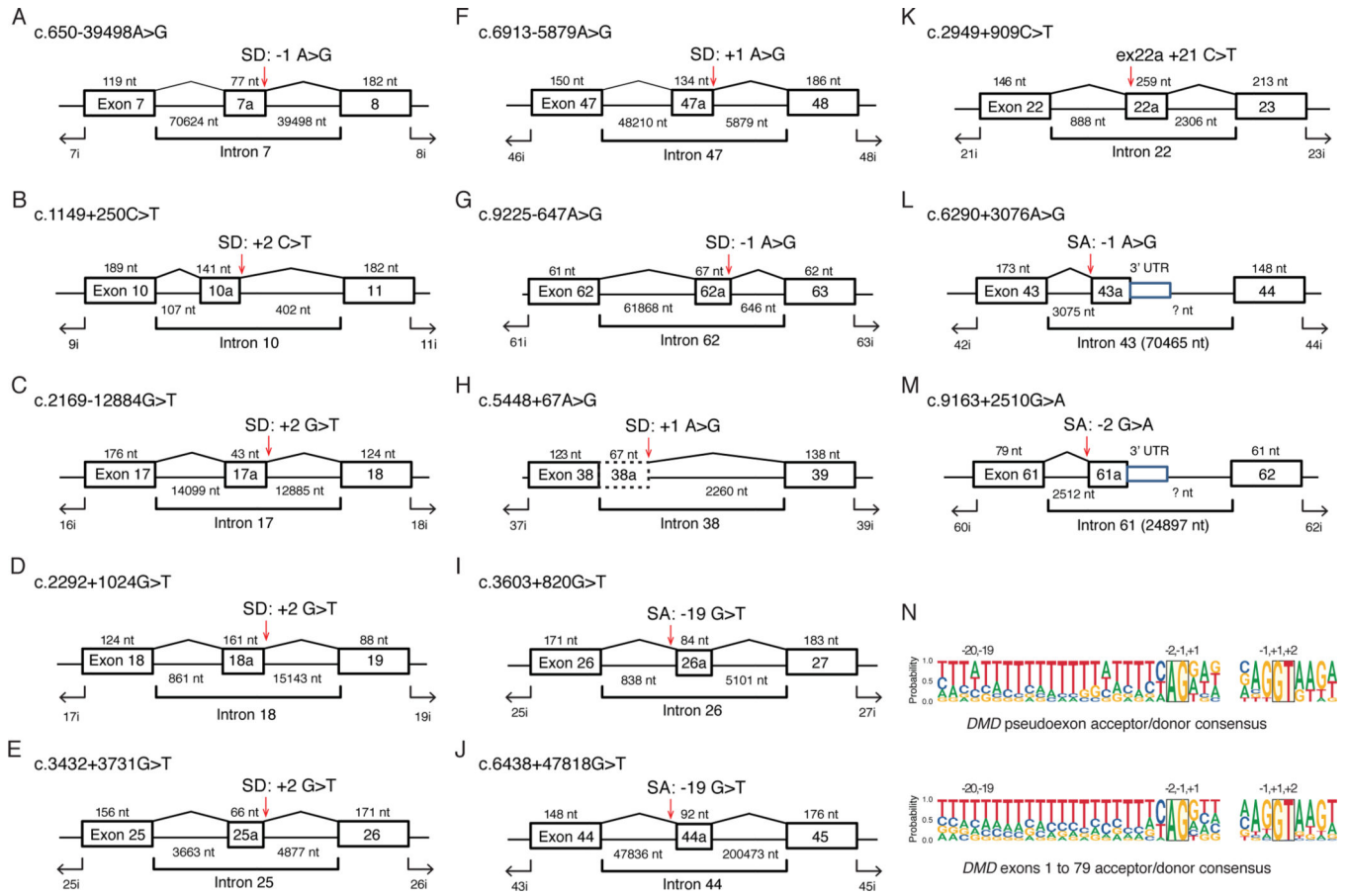
- Akhtar MN, Bukhari SA, Fazal Z, Qamar R, & Shahmuradov IA (2010). POLYAR, a new computer program for prediction of poly(A) sites in human sequences. *BMC Genomics*, 11, 646. doi:10.1186/1471-2164-11-646 [PubMed: 21092114]
- Ankala A, Kohn JN, Hegde A, Meka A, Ephrem CL, Askree SH, ... Hegde MR. (2012). Aberrant firing of replication origins potentially explains intragenic nonrecurrent rearrangements within genes, including the human DMD gene. *Genome Res*, 22(1), 25–34. doi:10.1101/gr.123463.111 [PubMed: 22090376]
- Anthony K, Arechavala-Gomez V, Ricotti V, Torelli S, Feng L, Janghra N, ... Muntoni F. (2014). Biochemical characterization of patients with in-frame or out-of-frame DMD deletions pertinent to exon 44 or 45 skipping. *JAMA Neurol*, 71(1), 32–40. doi:10.1001/jamaneurol.2013.4908 [PubMed: 24217213]
- Baskin B, Gibson WT, & Ray PN (2011). Duchenne muscular dystrophy caused by a complex rearrangement between intron 43 of the DMD gene and chromosome 4. *Neuromuscul Disord*, 21(3), 178–182. doi:10.1016/j.nmd.2010.11.008 [PubMed: 21134752]
- Beroud C, Hamroun D, Collod-Beroud G, Boileau C, Soussi T, & Claustres M. (2005). UMD (Universal Mutation Database): 2005 update. *Hum Mutat*, 26(3), 184–191. doi:10.1002/humu.20210 [PubMed: 16086365]
- Bladen CL, Salgado D, Monges S, Foncuberta ME, Kekou K, Kosma K, ... Lochmuller H. (2015). The TREAT-NMD DMD Database: analysis of more than 7,000 Duchenne muscular dystrophy mutations. *Hum Mutat*, 36(4), 395–402. doi:10.1002/humu.22758 [PubMed: 25604253]
- Blazquez L, Aiastui A, Goicoechea M, Martins de Araujo M, Avril A, Beley C, ... Lopez de Munain A. (2013). In vitro correction of a pseudoexon-generating deep intronic mutation in LGMD2A by antisense oligonucleotides and modified small nuclear RNAs. *Hum Mutat*, 34(10), 1387–1395. doi:10.1002/humu.22379 [PubMed: 23864287]
- Bolduc V, Foley AR, Solomon-Degefa H, Sarathy A, Donkervoort S, Hu Y, ... Bonnemann CG. (2019). A recurrent COL6A1 pseudoexon insertion causes muscular dystrophy and is effectively targeted by splice-correction therapies. *JCI Insight*, 4(6). doi:10.1172/jci.insight.124403
- Bouge AL, Murauer E, Beyne E, Miro J, Varilh J, Taulan M, ... Tuffery-Giraud S. (2017). Targeted RNA-Seq profiling of splicing pattern in the DMD gene: exons are mostly constitutively spliced in human skeletal muscle. *Sci Rep*, 7, 39094. doi:10.1038/srep39094 [PubMed: 28045018]
- Bovolenta M, Neri M, Fini S, Fabris M, TrabANELLI C, Venturoli A, ... Ferlini A. (2008). A novel custom high density-comparative genomic hybridization array detects common rearrangements as well as deep intronic mutations in dystrophinopathies. *BMC Genomics*, 9, 572. doi:10.1186/1471-2164-9-572 [PubMed: 19040728]
- Cagliani R, Sironi M, Ciafaloni E, Bardoni A, Fortunato F, Prella A, ... Comi GP. (2004). An intragenic deletion/inversion event in the DMD gene determines a novel exon creation and results in a BMD phenotype. *Hum Genet*, 115(1), 13–18. doi:10.1007/s00439-004-1118-6 [PubMed: 15118904]
- Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, Foley AR, ... MacArthur DG. (2017). Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med*, 9(386). doi:10.1126/scitranslmed.aal5209
- de Feraudy Y, Ben Yaou R, Wahbi K, Stalens C, Stantzou A, Laugel V, ... Amthor H. (2021). Very Low Residual Dystrophin Quantity Is Associated with Milder Dystrophinopathy. *Ann Neurol*, 89(2), 280–292. doi:10.1002/ana.25951 [PubMed: 33159473]
- Dent KM, Dunn DM, von Niederhausern AC, Aoyagi AT, Kerr L, Bromberg MB, ... Flanigan KM. (2005). Improved molecular diagnosis of dystrophinopathies in an unselected clinical cohort. *Am J Med Genet A*, 134(3), 295–298. doi:10.1002/ajmg.a.30617 [PubMed: 15723292]
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, ... Gingeras TR. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21. doi:10.1093/bioinformatics/bts635 [PubMed: 23104886]
- Dominov JA, Uyan O, Sapp PC, McKenna-Yasek D, Nallamilli BR, Hegde M, & Brown RH Jr. (2014). A novel dysferlin mutant pseudoexon bypassed with antisense oligonucleotides. *Ann Clin Transl Neurol*, 1(9), 703–720. doi:10.1002/acn3.96 [PubMed: 25493284]



- Dye MJ, & Proudfoot NJ (1999). Terminal exon definition occurs cotranscriptionally and promotes termination of RNA polymerase II. *Mol Cell*, 3(3), 371–378. doi:10.1016/s1097-2765(00)80464-5 [PubMed: 10198639]
- Flanigan KM (2014). Duchenne and Becker muscular dystrophies. *Neurol Clin*, 32(3), 671–688, viii. doi:10.1016/j.ncl.2014.05.002 [PubMed: 25037084]
- Flanigan KM, Dunn D, Larsen CA, Medne L, Bonnemann CB, & Weiss RB (2011). Becker muscular dystrophy due to an inversion of exons 23 and 24 of the DMD gene. *Muscle Nerve*, 44(5), 822–825. doi:10.1002/mus.22226 [PubMed: 22006698]
- Flanigan KM, Dunn DM, von Niederhausern A, Soltanzadeh P, Gappmaier E, Howard MT, ... Weiss RB. (2009). Mutational spectrum of DMD mutations in dystrophinopathy patients: application of modern diagnostic techniques to a large cohort. *Hum Mutat*, 30(12), 1657–1666. doi:10.1002/humu.21114 [PubMed: 19937601]
- Garcia-Rodriguez R, Hiller M, Jimenez-Gracia L, van der Pal Z, Balog J, Adamzek K, ... Spitali P. (2020). Premature termination codons in the DMD gene cause reduced local mRNA synthesis. *Proc Natl Acad Sci U S A*, 117(28), 16456–16464. doi:10.1073/pnas.1910456117 [PubMed: 32616572]
- Garrido-Martin D, Palumbo E, Guigo R, & Breschi A. (2018). ggsashimi: Sashimi plot revised for browser- and annotation-independent splicing visualization. *PLoS Comput Biol*, 14(8), e1006360. doi:10.1371/journal.pcbi.1006360
- Gazzoli I, Pulyakhina I, Verwey NE, Ariyurek Y, Laros JF, t Hoen PA, & Aartsma-Rus A. (2016). Non-sequential and multi-step splicing of the dystrophin transcript. *RNA Biol*, 13(3), 290–305. doi:10.1080/15476286.2015.1125074 [PubMed: 26670121]
- Ginsberg MR, McCarty AJ, Lacomis D, & Abdel-Hamid HZ (2018). Duchenne muscular dystrophy caused by a novel deep intronic DMD mutation. *Muscle Nerve*, 57(6), E136–E138. doi:10.1002/mus.26073 [PubMed: 29346715]
- Goncalves A, Oliveira J, Coelho T, Taipa R, Melo-Pires M, Sousa M, & Santos R. (2017). Exonization of an Intronic LINE-1 Element Causing Becker Muscular Dystrophy as a Novel Mutational Mechanism in Dystrophin Gene. *Genes (Basel)*, 8(10). doi:10.3390/genes8100253
- Gonorazky H, Liang M, Cummings B, Lek M, Micallef J, Hawkins C, ... Dowling JJ. (2016). RNAseq analysis for the diagnosis of muscular dystrophy. *Ann Clin Transl Neurol*, 3(1), 55–60. doi:10.1002/acn3.267 [PubMed: 26783550]
- Greer K, Mizzi K, Rice E, Kuster L, Barrero RA, Bellgard MI, ... Fletcher S. (2015). Pseudoexon activation increases phenotype severity in a Becker muscular dystrophy patient. *Mol Genet Genomic Med*, 3(4), 320–326. doi:10.1002/mgg3.144 [PubMed: 26247048]
- Gurvich OL, Tuohy TM, Howard MT, Finkel RS, Medne L, Anderson CB, ... Flanigan KM. (2008). DMD pseudoexon mutations: splicing efficiency, phenotype, and potential therapy. *Ann Neurol*, 63(1), 81–89. doi:10.1002/ana.21290 [PubMed: 18059005]
- Ishmukhametova A, Khau Van Kien P, Mechin D, Thorel D, Vincent MC, Rivier F, ... Tuffery-Giraud S. (2012). Comprehensive oligonucleotide array-comparative genomic hybridization analysis: new insights into the molecular pathology of the DMD gene. *Eur J Hum Genet*, 20(10), 1096–1100. doi:10.1038/ejhg.2012.51 [PubMed: 22510846]
- Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, ... Farh KK. (2019). Predicting Splicing from Primary Sequence with Deep Learning. *Cell*, 176(3), 535–548 e524. doi:10.1016/j.cell.2018.12.015 [PubMed: 30661751]
- Jones HF, Bryen SJ, Waddell LB, Bournazos A, Davis M, Farrar MA, ... Cooper S. (2019). Importance of muscle biopsy to establish pathogenicity of DMD missense and splice variants. *Neuromuscul Disord*, 29(12), 913–919. doi:10.1016/j.nmd.2019.09.013 [PubMed: 31706698]
- Keegan NP (2020). Pseudoexons of the DMD Gene. *J Neuromuscul Dis*, 7(2), 77–95. doi:10.3233/JND-190431 [PubMed: 32176650]
- Khelifi MM, Ishmukhametova A, Khau Van Kien P, Thorel D, Mechin D, Perelman S, ... Tuffery-Giraud S. (2011). Pure intronic rearrangements leading to aberrant pseudoexon inclusion in dystrophinopathy: a new class of mutations? *Hum Mutat*, 32(4), 467–475. doi:10.1002/humu.21471 [PubMed: 21305657]

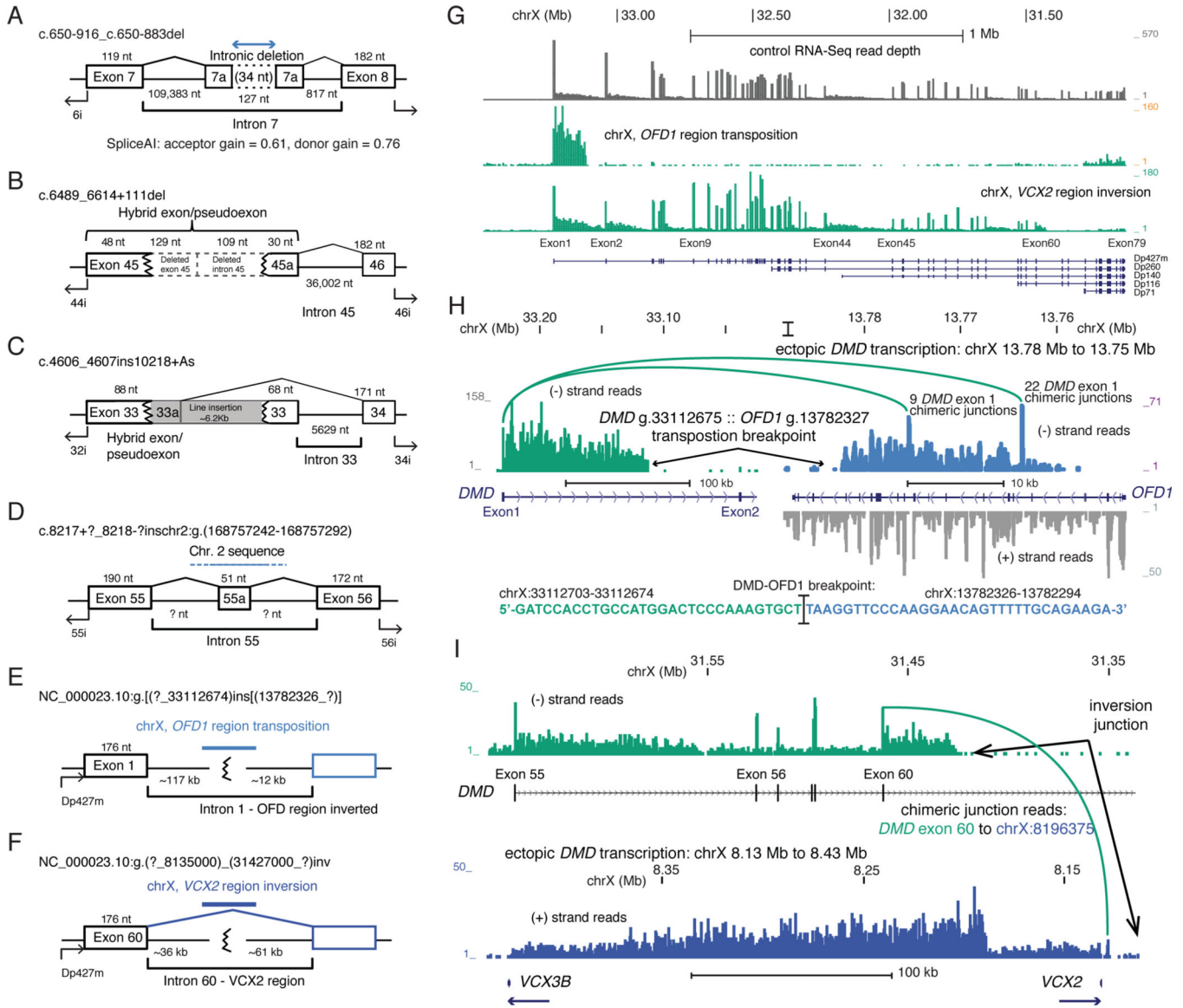
- Kinane TB, Mayer OH, Duda PW, Lowes LP, Moody SL, & Mendell JR (2018). Long-Term Pulmonary Function in Duchenne Muscular Dystrophy: Comparison of Eteplirsen-Treated Patients to Natural History. *J Neuromuscul Dis*, 5(1), 47–58. doi:10.3233/JND-170272 [PubMed: 29278896]
- Lai F, Damle SS, Ling KK, & Rigo F. (2020). Directed RNase H Cleavage of Nascent Transcripts Causes Transcription Termination. *Mol Cell*, 77(5), 1032–1043 e1034. doi:10.1016/j.molcel.2019.12.029 [PubMed: 31924447]
- Liao Y, Smyth GK, & Shi W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7), 923–930. doi:10.1093/bioinformatics/btt656 [PubMed: 24227677]
- Madden HR, Fletcher S, Davis MR, & Wilton SD (2009). Characterization of a complex Duchenne muscular dystrophy-causing dystrophin gene inversion and restoration of the reading frame by induced exon skipping. *Hum Mutat*, 30(1), 22–28. doi:10.1002/humu.20806 [PubMed: 18570328]
- Magri F, Del Bo R, D'Angelo MG, Govoni A, Ghezzi S, Gandossini S, ... Comi GP. (2011). Clinical and molecular characterization of a cohort of patients with novel nucleotide alterations of the Dystrophin gene detected by direct sequencing. *BMC Med Genet*, 12, 37. doi:10.1186/1471-2350-12-37 [PubMed: 21396098]
- Mendell JR, Goemans N, Lowes LP, Alfano LN, Berry K, Shao J, ... Telethon Foundation DMDIN. (2016). Longitudinal effect of eteplirsen versus historical control on ambulation in Duchenne muscular dystrophy. *Ann Neurol*, 79(2), 257–271. doi:10.1002/ana.24555 [PubMed: 26573217]
- Mendell JR, Rodino-Klapac LR, Sahenk Z, Roush K, Bird L, Lowes LP, ... Eteplirsen Study G. (2013). Eteplirsen for the treatment of Duchenne muscular dystrophy. *Ann Neurol*, 74(5), 637–647. doi:10.1002/ana.23982 [PubMed: 23907995]
- Mendell JR, Shilling C, Leslie ND, Flanigan KM, al-Dahhak R, Gastier-Foster J, ... Weiss RB. (2012). Evidence-based path to newborn screening for Duchenne muscular dystrophy. *Ann Neurol*, 71(3), 304–313. doi:10.1002/ana.23528 [PubMed: 22451200]
- Mitsui J, Takahashi Y, Goto J, Tomiyama H, Ishikawa S, Yoshino H, ... Tsuji S. (2010). Mechanisms of genomic instabilities underlying two common fragile-site-associated loci, PARK2 and DMD, in germ cell and cancer cell lines. *Am J Hum Genet*, 87(1), 75–89. doi:10.1016/j.ajhg.2010.06.006 [PubMed: 20598272]
- Monaco AP, Bertelson CJ, Liechti Gallati S, Moser H, & Kunkel LM (1988). An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. *Genomics*, 2(1), 90–95. [PubMed: 3384440]
- Muntoni F, Torelli S, & Ferlini A. (2003). Dystrophin and mutations: one gene, several proteins, multiple phenotypes. *Lancet Neurol*, 2(12), 731–740. [PubMed: 14636778]
- Oshima J, Magner DB, Lee JA, Breman AM, Schmitt ES, White LD, ... del Gaudio D. (2009). Regional genomic instability predisposes to complex dystrophin gene rearrangements. *Hum Genet*, 126(3), 411–423. doi:10.1007/s00439-009-0679-9 [PubMed: 19449031]
- Rendu J, Brocard J, Denarier E, Monnier N, Pietri-Rouxel F, Beley C, ... Marty I. (2013). Exon skipping as a therapeutic strategy applied to an RYR1 mutation with pseudo-exon inclusion causing a severe core myopathy. *Hum Gene Ther*, 24(7), 702–713. doi:10.1089/hum.2013.052 [PubMed: 23805838]
- Romitti PA, Zhu Y, Puzhankara S, James KA, Nabukera SK, Zamba GK, ... STARnet MD. (2015). Prevalence of Duchenne and Becker muscular dystrophies in the United States. *Pediatrics*, 135(3), 513–521. doi:10.1542/peds.2014-2044 [PubMed: 25687144]
- Sibley CR, Emmett W, Blazquez L, Faro A, Haberman N, Briese M, ... Ule J. (2015). Recursive splicing in long vertebrate genes. *Nature*, 521(7552), 371–375. doi:10.1038/nature14466 [PubMed: 25970246]
- Smith BF, Yue Y, Woods PR, Kornegay JN, Shin JH, Williams RR, & Duan D. (2011). An intronic LINE-1 element insertion in the dystrophin gene aborts dystrophin expression and results in Duchenne-like muscular dystrophy in the corgi breed. *Lab Invest*, 91(2), 216–231. doi:10.1038/labinvest.2010.146 [PubMed: 20714321]
- So BR, Di C, Cai Z, Venters CC, Guo J, Oh JM, ... Dreyfuss G. (2019). A Complex of U1 snRNP with Cleavage and Polyadenylation Factors Controls Telescripting, Regulating mRNA Transcription

- in Human Cells. *Mol Cell*, 76(4), 590–599 e594. doi:10.1016/j.molcel.2019.08.007 [PubMed: 31522989]
- Takeshima Y, Yagi M, Okizuka Y, Awano H, Zhang Z, Yamauchi Y, ... Matsuo M. (2010). Mutation spectrum of the dystrophin gene in 442 Duchenne/Becker muscular dystrophy cases from one Japanese referral center. *J Hum Genet*, 55(6), 379–388. doi:10.1038/jhg.2010.49 [PubMed: 20485447]
- Tennyson CN, Shi Q, & Worton RG (1996). Stability of the human dystrophin transcript in muscle. *Nucleic Acids Res*, 24(15), 3059–3064. doi:10.1093/nar/24.15.3059 [PubMed: 8760894]
- Treisman R, Orkin SH, & Maniatis T. (1983). Specific transcription and RNA splicing defects in five cloned beta-thalassaemia genes. *Nature*, 302(5909), 591–596. doi:10.1038/302591a0 [PubMed: 6188062]
- Tuffery-Giraud S, Beroud C, Leturcq F, Yaou RB, Hamroun D, Michel-Calemard L, ... Claustres M. (2009). Genotype-phenotype analysis in 2,405 patients with a dystrophinopathy using the UMD-DMD database: a model of nationwide knowledgebase. *Hum Mutat*, 30(6), 934–945. doi:10.1002/humu.20976 [PubMed: 19367636]
- Tuffery-Giraud S, Saquet C, Chambert S, & Claustres M. (2003). Pseudoexon activation in the DMD gene as a novel mechanism for Becker muscular dystrophy. *Hum Mutat*, 21(6), 608–614. doi:10.1002/humu.10214 [PubMed: 12754707]
- Waddell LB, Bryen SJ, Cummings BB, Bournazos A, Evesson FJ, Joshi H, ... Cooper ST. (2021). WGS and RNA Studies Diagnose Noncoding DMD Variants in Males With High Creatine Kinase. *Neurol Genet*, 7(1), e554. doi:10.1212/NXG.0000000000000554 [PubMed: 33977140]
- Waldrop MA, Gumienny F, El Husayni S, Frank DE, Weiss RB, & Flanigan KM (2018). Low-level dystrophin expression attenuating the dystrophinopathy phenotype. *Neuromuscul Disord*, 28(2), 116–121. doi:10.1016/j.nmd.2017.11.007 [PubMed: 29305136]
- Wang R, Zheng D, Wei L, Ding Q, & Tian B. (2019). Regulation of Intronic Polyadenylation by PCF11 Impacts mRNA Expression of Long Genes. *Cell Rep*, 26(10), 2766–2778 e2766. doi:10.1016/j.celrep.2019.02.049 [PubMed: 30840896]
- Wein N, Vulin A, Falzarano MS, Szigyarto CA, Maiti B, Findlay A, ... Flanigan KM. (2014). Translation from a DMD exon 5 IRES results in a functional dystrophin isoform that attenuates dystrophinopathy in humans and mice. *Nat Med*, 20(9), 992–1000. doi:10.1038/nm.3628 [PubMed: 25108525]
- Wimmer K, Schamschula E, Wernstedt A, Traunfellner P, Amberger A, Zschocke J, ... Messiaen L. (2020). AG-exclusion zone revisited: Lessons to learn from 91 intronic NF1 3' splice site mutations outside the canonical AG-dinucleotides. *Hum Mutat*, 41(6), 1145–1156. doi:10.1002/humu.24005 [PubMed: 32126153]
- Yan J, Feng J, & Buzin CH (2004). Three-tiered noninvasive diagnosis in 96% of patients with Duchenne muscular dystrophy (DMD). *Hum Mutat*, 23(2), 203–204. [PubMed: 14722924]
- Zaum AK, Stuve B, Gehrig A, Kolbel H, Schara U, Kress W, & Rost S. (2017). Deep intronic variants introduce DMD pseudoexon in patient with muscular dystrophy. *Neuromuscul Disord*, 27(7), 631–634. doi:10.1016/j.nmd.2017.04.003 [PubMed: 28495050]
- Zhao Z, Xu Q, Wei R, Wang W, Ding D, Yang Y, ... Ni T. (2021). Cancer-associated dynamics and potential regulators of intronic polyadenylation revealed by IPAfinder using standard RNA-seq data. *Genome Res*, 31(11), 2095–2106. doi:10.1101/gr.271627.120 [PubMed: 34475268]



**Figure 1. Deep Intronic *DMD* Point Mutations that Activate Pseudoexons**

(A-H) Individual exon/intron diagrams for mutations that create cryptic splice acceptor (SA) or splice donor (SD) sites used for pseudoexon inclusion. The coding DNA (c.) nomenclature is based on NM\_004006.2, and exon/intron sizes (not drawn to scale) are shown in nucleotides (nt) with mutation locations indicated with red arrows. (I,J) Mutations that inactivate AG dinucleotides in the polypyrimidine tract. (K) Exonic splice enhancer mutation +21 nucleotides from the 5' splice junction. (L,M) Mutations activating cryptic SA sites associated with pseudo 3'-terminal exons. (N) Sequence logos showing the relative nucleotide proportions (probability) at each position of the pseudoexon SA and SD sites (upper panel) versus constitutive splice sites for *DMD* exons 1 to 79 of NM\_004006.2.



**Figure 2. Type 2 and 3 DMD mutations**

Exon/intron diagrams for intronic deletions, insertions and chromosome level rearrangements. Mutations that (A) activate pseudoexon inclusion by a 34 nt. intronic deletion, (B) cryptic splice donor activation by a 238 bp. Deletion, or (C) a ~6.2 kb L1 retrotransposon insertion. Complex mutations due to (D) an inter-chromosomal transposition of a chr2 segment, (E) intra-chromosomal transposition of the chrX *OFD1* region, and (F) a large, 23 Mb inversion between the *VCX2-3* region at chrX:8.2 Mb and *DMD* intron 60 at chrX:31.4 Mb. (G) RNA-Seq read depth across the *DMD* region for the (E) and (F) mutations. (H) RNA-Seq coverage and exon junction reads that link *DMD* exon 1 with the *OFD1* region, and the genomic sequence of the confirmed *DMD* intron 1 to *OFD1* region junction. Strand-specific coverage is shown in the *OFD1* regions, ectopic transcription on the minus strand shown in brown. (I) RNA-Seq coverage showing transcription from

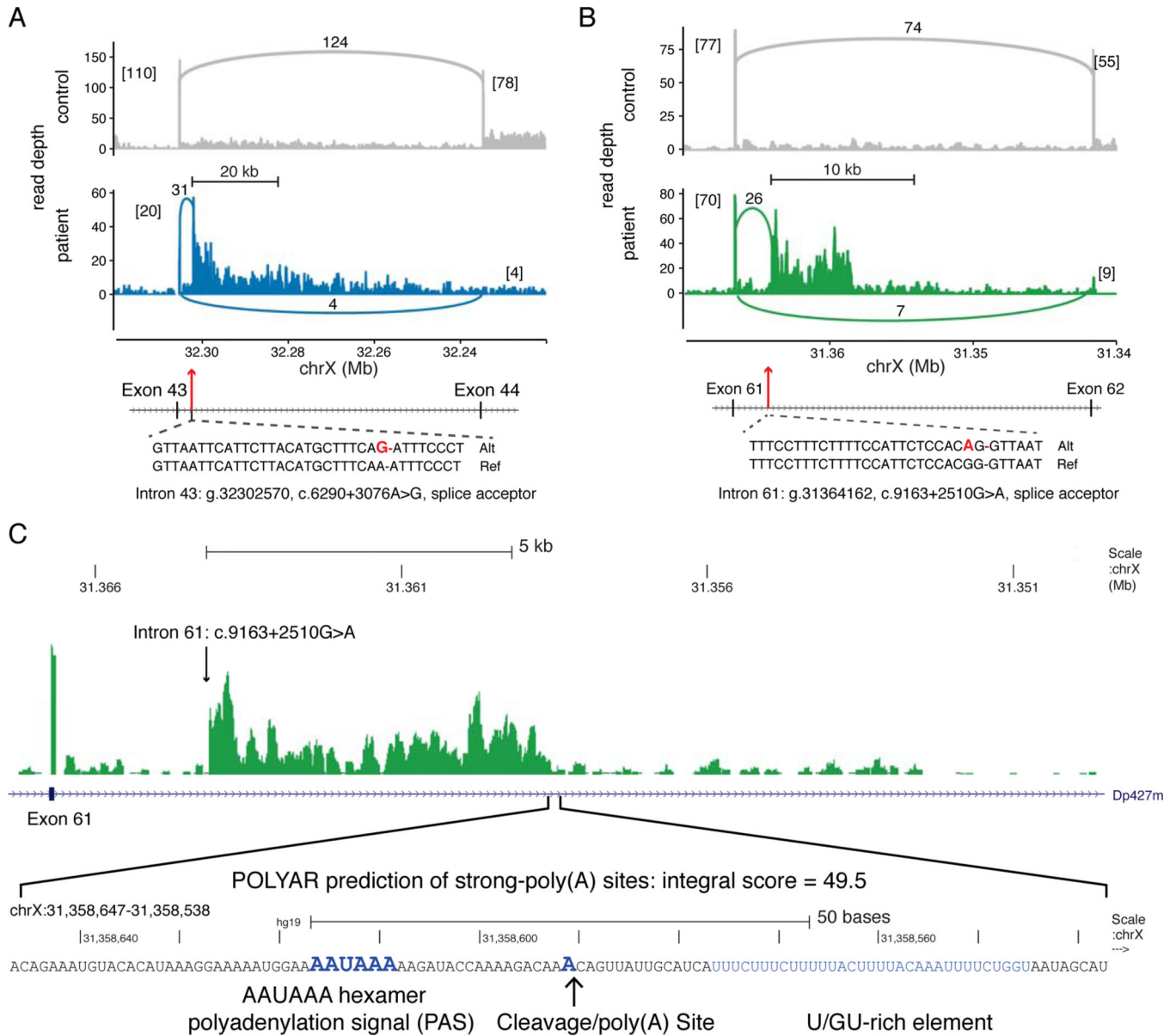
the *DMD* region and exon 60 junction reads supporting the inversion breakpoint junction (green), and ectopic transcription through the *VCX2*–*VCX3B* region (blue).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 3. Type 1C Intronic DMD Mutations Activate Intronic Splice Acceptor Sites**

(A) Coverage plots of stranded, total RNA-Seq reads mapped to the *DMD* exon 43 – intron 43 – exon 44 region (hg19, chrX:32.32–32.22 Mb). The lower panel (blue) displays mapped reads from the intron 43 c.6290+3076A>G patient, compared to reads from a control muscle biopsy (upper panel, gray). Exon and pseudoexon junction read counts are shown as numbers alongside connecting arcs. Numbers in brackets are flanking exon junction read counts. The cryptic splice acceptor site activating mutation (bold red) is shown relative to the reference sequence. (B) Coverage plots for mapped reads in the *DMD* exon 61 – intron 61 – exon 62 region (chrX:31.37–31.34 Mb) from the intron 61 c.9163+2510G>A patient (lower panel, green) versus a control biopsy (upper panel, gray). (C) The read depth for the intron 61 c.9163+2510G>A patient is zoomed to the intron 60 – exon 61 – intron 61 region (chrX:31,349,600–31,367,400) and then to the 100 nucleotide region (chrX:31,358,538–

31,358,647) with an abrupt decrease in read depth coinciding with a POLYAR predicted strong cleavage and polyadenylation site.

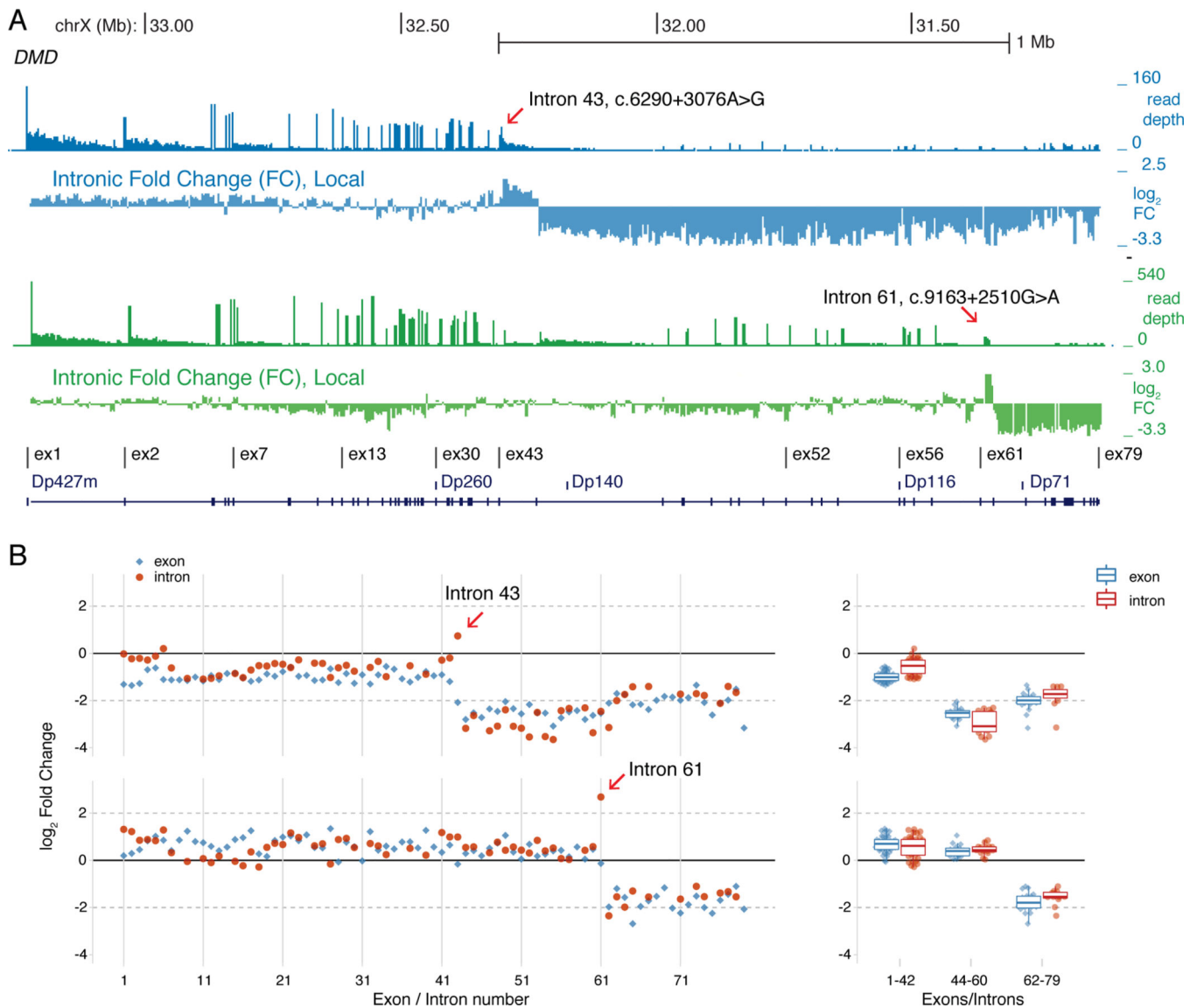
Author Manuscript

Author Manuscript

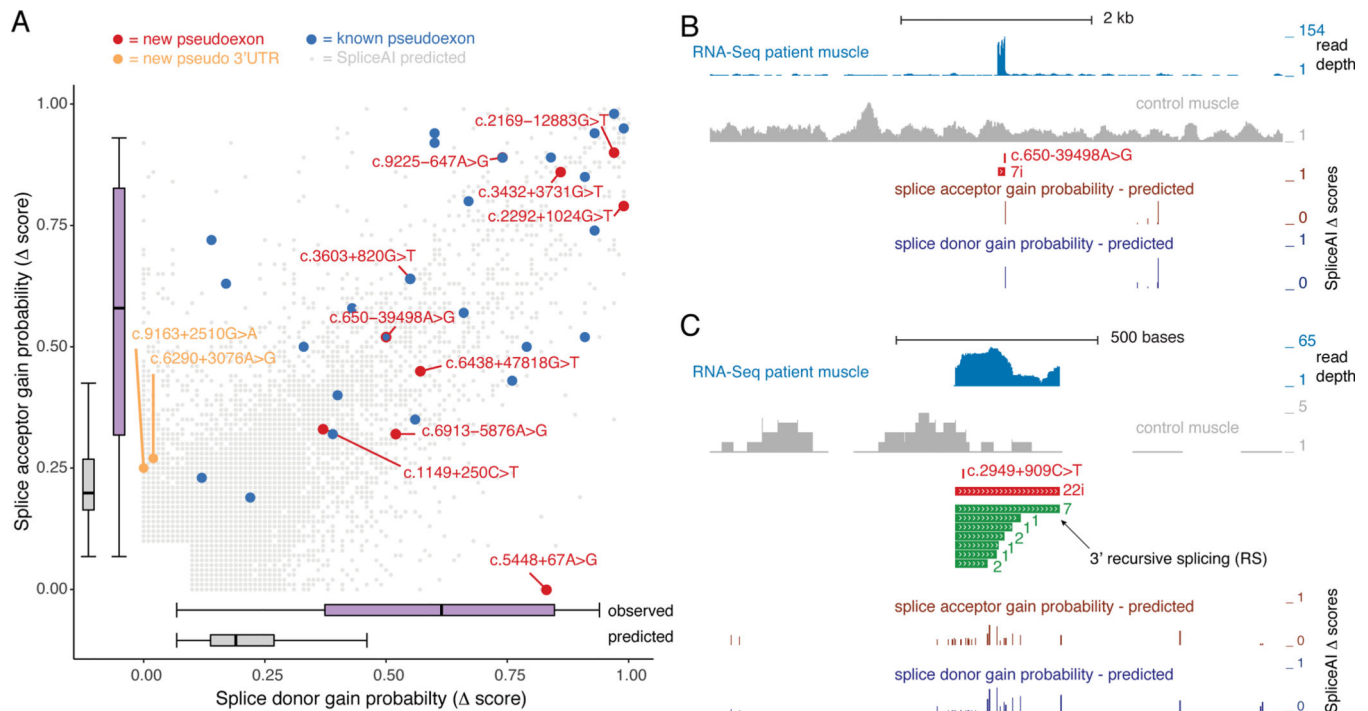
Author Manuscript

Author Manuscript





**Figure 4. Type 1C Intronic *DMD* Mutations Cause Premature RNA Transcription Termination**  
 (A) Read depth and differential intron coverage tracks across the *DMD* gene (chrX:33.27–31.11 Mb) with the intron 43 c.6290+3076A>G patient browser tracks shown in blue (upper) and the intron 61 c.9163+2510G>A patient in green (lower). The ‘Intronic Fold Change, Local’ track plots the  $\log_2$  ratio of patient versus an unaffected 9-year-old male from read counts summarized in 5 kb non-overlapping intronic bins and normalized to the total number of *DMD* intronic reads in each sample. (B) Globally normalized RNA-seq data for *DMD* for patients with the indicated *DMD* mutation, with the left panels plotting 5’ to 3’ the intron (red) and exon (blue)  $\log_2$  fold change differences from compared to values from an unaffected 9-year-old male. Right panels are box plots of  $\log_2$  fold changes for exons/introns grouped from 1 to 44, 44–60, and 62 to 79.



**Figure 5. Landscape of Observed and Predicted *DMD* Pseudoexon Mutations**

A) SpliceAI-10k predicted strength gains of cryptic splice sites caused by SNVs in *DMD* intronic regions (excluding 50 nts. flanking each exon) are shown for donor and acceptor sites. The probability of the SNV causing cryptic splice site gain (D score) ranges from 0 to 1, and only sites with a D score  $\geq 0.1$  were included. Point mutations described in Figure 1A–J are shown in red, pseudo 3'-terminal exons mutations from Figure 1L–M are shown in orange, and point mutations found in the literature (Table S2) are shown in blue. Box plots summarizing the D score ranges for the acceptor and donor site scores are shown separately for observed sites (new + known, purple) and predicted sites (gray). (B) Detailed view of intron 7 pseudoexon mutation (see Figure 1A) with RNA-Seq coverage from patient muscle. Locations of the SpliceAI-10k predictions for splice acceptor (brown) or donor (blue) gain are shown for all scores  $\geq 0.1$ . (C) Detailed view of intron 22 pseudoexon mutation (see Figure 1K) with the observed pseudoexon shown in red and overlapping SpliceAI predicted pseudoexons in green. The arrow indicates the overlap of a 3'RS site.

General Classification of Splice-Altering Intronic Mutations

Table 1:

Intronic variation type	Subtype	Effect on splicing
1 (point mutations)	1a	Point mutations creating a consensus splice donor or acceptor site
	1b	Point mutations creating an exonic splice enhancer
	1c	Point mutations creating an acceptor site resulting in the inclusion of a cryptic 3' terminal exon
2 (substitutions, deletions or insertions)		Variants resulting in inclusion of novel pseudoexons without alteration of clearly defined splicing motifs
3 (translocations)		Translocations (or apparent translocations) of sequences from other chromosomal loci into introns, altering splicing patterns

Genotypes and Phenotypes of DMD/BMD Patients

Table 2:

Figure Key	Subtype	intron	HGVS Genomic_GRCCh37 (NC_000223.10)	cDNA (NM_004006.2)	RNA	Protein	Frame as predicted by RNA sequencing	Clinical Diagnosis	Presenting Symptoms (age at onset)	Ambulation status (age)	Cardiac Status
Fig. 1A	1a	7	chrX:32756908T>C	c.650-39498A>G	r.649_650i ns650-39575_650-39499	p.Asp217Alafs2	OUT	DMD	Abnormal gait, frequent falls, difficulty climbing stairs, thigh pain (2y)	ambulant (4y)	Normal (4y)
Fig. 1B	1a	10	chrX:32662831G>A	c.1149+250 C>T	r.1149_1150ins1149+108_1149+248	p.Gly384LeufsTer3	OUT	BMD	Trouble with stairs (7y)	ambulant (23y)	Normal (23y)
Fig. 1C	1a	17	chrX:32549132C>A	c.2169-12883G>T	r.2168_2169i ns2168+14099_2168+14142	p.Leu723Ter	OUT	BMD	3y	ambulant (32y)	Normal (32y)
Fig. 1D	1a	18	chrX:32535101C>A	c.2292+1024 G>T	r.2292_2293i ns2292+862_2292+1022	p.Ala765AspfsTer23	OUT	DMD	Calf hypertrophy (3y)	wheelchair fulltime (12y)	Normal (13y)
Fig. 1E	1a	25	chrX:32477825C>A	c.3432+3731 G>T	r.3432_3433i ns3432+3664_3432+3729	p.Val1145IlefsTer18	OUT	BMD	Difficulty carrying buckets of water up stairs (56y)	ambulant (68y)	HTN, h/o MI, no CM
Fig. 1F	1a	47	chrX:31899369T>C	c.6913-5879A>G	r.6912_6913i ns6913-6013_6913-5880	p.Val2305Ter	OUT	BMD	pain in legs (3.5y)	ambulant (8y)	Normal
Fig. 1G	1a	62	chrX:31279780T>C	c.9225-647A>G	r.9224_9225ins9225-713_9225-647	p.Asn3075LysfsTer2	OUT	HyperCK/BMD	motor & cognitive delays (1y)	ambulant (21y)	EF 54%, partial RBBB
Fig. 1H	1a	38	chrX:32366456T>C	c.5448+67A>G	r.[5448_5449i ns5448+1_5448+67]	p.Met1816_Asn1817insTer4	OUT	DMD	Gross motor delay (10m)	wheelchair fulltime (12y)	EF 42% (20y)
Fig. 1I	1a	26	chrX:32187215C>A	c.3603+820G>T	r.3603_3604i ns3603+839_3603+923	p.Arg1202AspfsTer16	OUT	DMD	gross motor delay (18m)	wheelchair fulltime (9y)	Normal (11y)
Fig. 1J	1a	44	chrX:32471959C>A	c.6438+47818G>T	r.6438_6439i ns6438-47837_6438-47929	p.Lys2146ValfsTer1	OUT	BMD	incidental finding of elevated AST	ambulant (16 yo)	Normal 16 yo, EF61%
Fig. 1K	1b	22	chrX:32489372G>A	c.2949+909 C>T	r.[2949_2950ins2949+889_2949+1147]; [=]	p.Ala984ThrfsTer33	OUT	HyperCK/BMD	None	ambulant (25y)	Normal (20y)

Figure Key	Subtype	intron	HGVs_Genomic_GRCh37 (NC_000023.10)	cDNA (NM_004006.2)	RNA	Protein	Frame as predicted by RNA sequencing	Clinical Diagnosis	Presenting Symptoms (age at onset)	Ambulation status (age)	Cardiac Status
Fig. 1L	1c	43	chrX:32302570T>C	c.6290+3076 A>G	r.[6290_6291-67389_6291+?]; [=]	p.Thr3055SerfsTer1	OUT	BMD	mild proximal weakness, calf hypertrophy, CK 10,000s	ambulant (7y)	NR
Fig. 1M	1c	61	chrX:31364163C>T	c.9163+2510G>A	r.[9163_9164-22385_9164+?]; [=]	p.Gly2097GlyfsTer2	OUT	HyperCK/BMD	muscle pain after play (4y)	ambulant (8y)	Normal (7y)
Fig. 2A	2	7	chrX:32718293_32718326del	c.650-916_c.650-883del	r.649_650i ns650-944_650-818del650-916_650-883	p.Asp217GlyfsTer2	OUT	DMD	motor delay (6m)	wheelchair fulltime (11.75y)	Abnl FS 24.6%
Fig. 2B	2	45*	chrX:31986547_31986583del	c.6489_6614+111del	r.[6487_6613delins6613+110_6613+139]; [=]	p.Asp2163_Arg2205delins10 (KLNSTINQQQ)	IN	BMD	difficulty climbing stairs (40y)	ambulant (70s)	Normal
Fig. 2C	2	33*	g.32404494-32404495i ns[10218];[A]	c.4606_4607i ns10218+As	r.4606_4674delins95	p.Glu1536GlyfsTer19	OUT	DMD	Abnormal gait, difficulty climbing stairs (2y)	Wheelchair fulltime (9.9y)	Normal (16y)
Fig. 2D	3	55	-	c.8217_8218i ns[NR_131227.1:c.37-87]	r.8217_8218i ns[NR_131227.1:c.37-87]	p.Glu2739_Asp2740i ns17	IN	DMD	Gross motor delay (7y)	Wheelchair fulltime (9y)	Normal (12y)
Fig. 2E	3	1	g.(?_33112674)i ns[g.(13782326_?)]	-	-	-	OUT	DMD	lower limb weakness, abnormal gait, elevated CK (5y)	ambulant (9y)	Normal (9y)
Fig. 2F	3	60	g.8135000_31427000i nv	-	-	-	OUT	DMD	Gross motor delay (walked 21 months & cognitive (2y)	wheelchair fulltime (9y)	Normal (13y)

NR = not reported; y = year

Hum Mutat. Author manuscript; available in PMC 2023 April 01.