Review

Pseudoexons of the DMD Gene

Niall P. Keegan∗
Centre for Molecular Medicine and Innovative Therapeutics, Murdoch University and Perron Institute, Perth, Australia

Abstract. The DMD gene is the largest in the human genome, with a total intron content exceeding 2.2Mb. In the decades since DMD was discovered there have been numerous reported cases of pseudoexons (PEs) arising in the mature DMD transcripts of some individuals, either as the result of mutations or as low-frequency errors of the spliceosome. In this review, I collate from the literature 58 examples of DMD PEs and examine the diversity and commonalities of their features. In particular, I note the high frequency of PEs that arise from deep intronic SNVs and discuss a possible link between PEs induced by distal mutations and the regulation of recursive splicing.

Keywords: Cryptic Splice Sites, Muscular Dystrophy, Duchenne, DNA Mutational Analysis, RNA Splicing

CHARACTERISTICS OF THE DMD GENE

The DMD gene is the largest gene in the human genome. Situated on the p-arm of the X chromosome, DMD spans over 2.22Mb, more than 99% of which is intronic sequence, with the coding sequence of its largest isoform totalling 11,058 bases across 79 exons. Eight unique alternative promoters [1], alternatively spliced exons, and an alternative polyadenylation site [2] produce at least 17 DMD transcript variants [3], one or more of which are expressed and translated in all types of muscle as well as various other cell types throughout the body, including myoblasts, lymphocytes and retinal cells.

The Dp427 m transcript of DMD encodes the muscle isoform of dystrophin, the DMD protein. In XY individuals, who carry just a single DMD copy, mutations that fully disrupt the function of the DMD gene (resulting in functionless or absent dystrophin protein) give rise to Duchenne muscular dystrophy, while mutations that only partially diminish the gene’s function and/or quantity of product give rise to Becker Muscular Dystrophy.

THE MAJOR SPLICEOSOME

The vast majority of RNA splicing in humans is achieved via the major spliceosome, a ribonucleoprotein complex responsible for excising introns from pre-mRNA molecules [4]. In order for the spliceosome to process a transcript correctly it first must accurately recognize the transcript’s exon-intron boundaries. This recognition is achieved through a network of mechanisms, including sequence-specific interaction with conserved acceptor and donor splice site motifs in the RNA, silencer and enhancer binding motifs both proximal and distal to the splice junctions, and RNA secondary structure [5, 6]. Mutations to a gene that alter the interactions of these factors with its transcripts can lead to errors in the processing of those transcripts, such as the expansion, truncation or loss of canonical exons, or the initiation of pseudoexons (PEs) within its introns [7]. These incorrectly spliced transcripts may be degraded prior to translation or may be translated to less functional or even harmful protein isoforms, with deleterious consequences for the health of the patient.

Thirty-six of the 78 introns in DMD are more than ten times the human median intron length of 1334bp [4], and of these 36, three are more than 100 times the median size. A transcript of this size and complexity...
presents a unique challenge to the major spliceosome, and as a result is arguably more vulnerable to splicing errors such as pseudoexons.

**PSEUDOEXONS: ONE DEFINITION OF MANY**

In the literature, the terms “pseudoexon” and “cryptic exon” are often applied interchangeably and inconsistently in reference to a wide range of splicing errors. For the sake of clarity, I hereby define a pseudoexon as: Any continuous tract of a transcribed gene that: (1) does not overlap, adjoin or duplicate any sense-strand sequence of that gene’s canonical exons; (2) bears an acceptor splice site motif at its 5′ end and a donor splice site motif at its 3′ end; and, (3) via both these motifs, is spliced into a measurable proportion of the mature transcripts of that gene in at least one proband.

Though this definition of “pseudoexon” may not agree with every prior usage of the term, it includes the majority of prior use while excluding splicing events that are better described by other terms, such as cryptic splice sites and whole exon duplications.

While some PEs are observable as rare splicing events in normal individuals, the majority are created by mutations that give the PE site an exon-like profile, resulting in the spliceosome falsely recognizing it and splicing it into an increased proportion of transcripts. When PE-splicing levels are high compared to normal splicing, these inclusions are likely to bear negative consequences for the phenotype of the affected organism, as the majority of PEs will disrupt the transcript’s open reading frame and/or encode premature stop codons. Consequently, the resulting transcript, if it is not degraded by nonsense-mediated decay, will be translated to a non-functional or truncated protein. Even in cases where a pseudoexon preserves the reading frame and does not encode a premature stop codon, it is likely that the amino acids it encodes will disrupt the secondary structure of the protein and thereby abrogate its function.

Numerous DMD PEs have been reported over the last few decades, perhaps more than have been described for any other single gene. When considered as a body of research, these reports comprise a unique opportunity for generating new insights into the splicing of DMD and other large genes.

In this review, I catalog the characteristics of 58 reported DMD PEs. Where possible, I describe the origins of these rare splicing events and draw inferences from their common features.

**PSEUDOEXONS OF THE DMD GENE**

Following a thorough search of the literature, I compiled a catalog of all 58 known PEs of the DMD gene (Table 1). In order to consistently record highly similar PEs, I adopted the criteria of unique local sequence: I assigned a separate catalog entry to each PE that was unique in at least one nucleotide of its sequence or splice motifs (eg. PE09 vs. PE10, PE15 vs. PE16); and listed as single entries all PEs with locally identical sequence (eg. PEs 11 and 12).

**PSEUDOEXONS ARISING WITHOUT MUTATIONS**

Six DMD pseudoexons have been reported as low-frequency splicing events in normal cells lacking any known mutation: PEs 04, 07, 08, 11, 21 and 44. Interestingly, four of these PEs are of lengths that do not shift the reading frame of the transcript – 162bp for PE04, 357 bp for PE11, 66bp for PE21 and 84bp for PE44 – and of these four, only PEs 04 and 11 contain stop codons.

In addition to their splicing profile in normal cells, PEs 04, 07, 08 and 11 are also spliced at much higher frequencies in the cells of some patients with other DMD mutations. For PEs 08 and 11, this behavior was observed for the cells of only a single patient each (see below, subsection ‘Pseudoexons arising from duplications’). However, the behavior of PEs 04 and 07 – referred to in some prior literature as exons 1a [10–14] and 2a [16, 17] respectively – is somewhat more complex. Though the inclusion of PE04 in muscle cell DMD transcripts is rare [14], this PE is included in approximately 50% of DMD transcripts in lymphocytes [11], and is included at higher frequencies in both cell types as a result of a frame-shifting single nucleotide deletion in exon 5 [12] and, in a different proband, an exon 2 tandem duplication [13]. Given the frequency of its inclusion in mature DMD transcripts, especially in normal lymphocytes, it may be that PE04 would be better classified as a canonical, alternatively spliced exon rather than a pseudoexon. However, at the time of writing no functional role for PE04 has been conclusively determined [10]. Similarly, PE07 has also been observed as a predominant inclusion in the muscle DMD transcripts of multiple patients with deletions and duplications in the 5′
Table 1
Details of 58 known pseudoexons (PEs) arising from mutations within the human DMD gene. Bold, upper-case sequence denotes the PE itself, while regular lower-case sequence indicates flanking intron. Inverted sequence is italicized. Duplications and insertions are underlined. Single base transitions are shown in the form of “(X > Y),” where X is the reference nucleotide and Y is the mutant. Shapiro-Senapathy splice scores are also shown, before and after mutation where appropriate (see supplementary materials for splice score calculator). “RNA Source” lists the cell or tissue types where the PE was detected. Genomic co-ordinates pertain to human genome assembly accession GCA_000001405.27, chromosome X, except where otherwise stated. Transcript co-ordinates pertain to DMD reference sequence NG_012232.1 (DMD_v001)

| PE # | Intron | Sequence | Genomic Co-ordinates (chrX.) | Size (bp) | Acceptor Score | Donor Score | Mutation(s) | Orientation of Mutation to PE and Splice Sites | ORF | RNA Source | Reference |
|------|--------|----------|----------------------------|-----------|----------------|-------------|-------------|-----------------------------------------------|-----|------------|-----------|
| 01   | 1      | tctcttccttggtttggc(g>a)gC | 33174185-33174333 | 149 | 71.37 -> 87.94 | 72.72 | c.31+36947G>A | Proximal Frameshift | Skeletal muscle | [9] |
| 02   | 1      | etctatattttgataagTAACA | 33088244-33088348 | 105 | 62.92 | 81.59 | c.exon3, 6del | Distal (3') Premature stop codon | Heart, skeletal muscle | [10] |
| 03   | 1      | tttttctcttgcataagTAACA | 33085901-33085927 | 27 | 58.89 | 78.81 | c.exon3, 6del | Distal (3') Premature stop codon | Heart, skeletal muscle | [10] |
| 04   | 1      | ttctgctctctcattagCAAC | 33078186-33078347 | 162 | 96.38 | 88.39 | Multiple mutations, also observed in normal cells (see main text) | Distal (3') Premature stop codon | Lymphocytes | [10-14] |
| 05   | 2      | tttttcttctcttctcatt(tag)G | 33014502-33014547 | 46 | 68.17 -> 84.74 | 82.07 | c.93+5590T>A | Proximal Frameshift | Lymphocytes | [15] |

(Continued)
| PE # | Intron Sequence | Genomic Co-ordinates (chrX.) | Size (bp) | Acceptor Score | Donor Score | Mutation(s) | Orientation of Mutation to PE and Splice Sites | ORF | RNA Source | Reference |
|------|----------------|-----------------------------|-----------|----------------|-------------|-------------|------------------------------------|------|------------|----------|
| 06   | ttcctgttttctctacat (t>a) | 33014416-33014547 | 132       | 68.17 -> 84.74 | 77.7        | c.93+5590T>A | Proximal | ORF preserved | Lymphocytes, skeletal muscle | [15] |
| 07   | tgttaggttttttggagcaggtct | 32960208-32960347 | 140       | 100            | 77.84       | Multiple deletions and duplications [17]; also observed in normal cells [16] | Distal (3’) | Frameshift | Skeletal muscle | [16, 17] |
| 08   | agttctggttctgtagctagt | 32878883-32878980 | 98        | 72.84          | 76.65       | c.exon2dup; also observed in normal cells | Distal (3’) | Frameshift | Multiple tissue types | [13] |
| 09   | tgattgtaaactttctgtagc | 32863759-32863915 | 157       | 70.85          | 73.68       | c.exon8.11dup | Distal (3’) | Frameshift | Lymphocytes | [18] |
| 10   | tggagactctctctgtagacag | 32863759-32863911 | 154       | 79.35          | 73.68       | c.exon8.11dup | Distal (3’) | Frameshift | Lymphocytes | [18] |
| 11 | 3 | aatcttactcttgtgccccag | 32846622-32846978 | 357 | 81.06 | 79.88 | c.exon8.1 dup; also observed in normal cells | Distal (3') | Premature stop codon | Lymphocytes | [18] |
| 12 | 3 | gtaaaaatggagattctggt | 32846622-32846683 | 62 | 72.62 | 79.88 | c.280Adel | Distal (3') | Frameshift | Lymphocytes | [12, 18] |
| 13 | 4 | ttcctcttttgggggaaagt | 32823851-32823982 | 132 | 90.56 | 74.59 | c.265-463A>G | Proximal Premature stop codon | Skeletal muscle | [19] |
| 14 | 7 | atattctaatgtatctcag (A>C) | 32738792-32738868 | 77 | 79.91 | 56.78 | c.650-3957A>C; c.650-39498A>G | Proximal Frameshift | Skeletal muscle | [20] |
| 15 | 9 | ttatattttccttctatta (A>C) | 32650985-32651074 | 90 | 83.55 | 70.04 | c.961-5925A>C | Proximal Premature stop codon | Skeletal muscle | [21] |

(Continued)
| PE # | Intron | Sequence | Genomic Co-ordinates (chrX.) | Size (bp) | Acceptor Score | Donor Score | Mutation(s) | Orientation of Mutation to PE and Splice Sites | ORF | RNA Source | Reference |
|------|--------|----------|-----------------------------|----------|---------------|-------------|-------------|-----------------------------------------------|-----|------------|-----------|
| 16   | 9      | ttcattttctccatattaagAGATT  |
|      |        | GAPCATATGGGTAGTTTAAACAGCTGTTT |
|      |        | TGGAAACAAACACCTTTGTTT |
|      |        | AGTCCAGACTCAGGTCTTTT |
|      |        | GTGACCTTTTg(c>t)agttcatct |
|      |        | aatatttctt           |
|      |        | 32650985-32651074     | 90        | 83.55         | 70.28 -> 87.27 | c.961-583C>T | Proximal | Premature stop codon | Skeletal muscle | [9]       |
| 17   | 11     | cttttttctctctctacctaagACAGG |
|      |        | GTTGGAAGATAGCTCGGAAGGCC |
|      |        | ATTTTCTCTGCTAAGCC(del: |
|      |        | 11631bp)AAACCTTTTCTACAAGA |
|      |        | ATAGTAAAAGCGCTTTTCTCAATCT |
|      |        | TAAAGAAAAAGATGTTAATGAGCAAA |
|      |        | TGAAGTCATCGCAGTAAACAAA |
|      |        | TCACTTGGTGAAGTTggttt |
|      |        | actatttctgccat         |
|      |        | 32641751-32641799;32630010-32630119 |
|      |        | 159        | 73.92         | 79.54         | c.1331+2382,1331+14010del | Proximal | Premature stop codon | Multiple tissue types | [22 - 25] |
| 18   | 11     | cttttttttttcccccaagTGCTC |
|      |        | ATTGAGCTCTGAAATAGATGACCA |
|      |        | TATGAGATGGTGAAGAGAAAAAGG |
|      |        | AGTGGCTGCAACATTGTGCTTTA |
|      |        | TTTACTCCAAAGGCTAGTAGGTTT |
|      |        | GCGTAGAATCGATATGAAATGA |
|      |        | GAGAACAACAAATAATAGTATTATC |
|      |        | AGgtaattccaatctgggatatt |
|      |        | 32638732-32638888     | 157        | 79.76         | 87.83         | c.1336,1337del | Distal (3′) | Frameshift | Lymphocytes | [26]       |
| 19   | 11     | agtttattttctctctccacccagGCTGG |
|      |        | AGTGGAGTTGGTGATGCTACGCTCA |
|      |        | CTGCAAACCTTCTGCCCCTGGGTTTC |
|      |        | AGTGGATTTCTCTGCTCGAGCCTCC |
|      |        | (c>g)tgagtcgggtattataaaa |
|      |        | ttttttttccacctctgctatttgAGAA |
|      |        | GAGGCTATCTTCAAACGTGAAAAAGG |
|      |        | CCTACCTCTGGGCTGTTCTCACAG |
|      |        | CAGCATCTGACTGATGATTACAG |
|      |        | CACCCTGCTTCCATCTAGCTGATAT |
|      |        | CCAAGATTGCAAAAAACTCAGACCC |
|      |        | AGgtcttttaatattaggttcctta |
|      |        | 32626363-32626441     | 79        | 90.08         | 56.56 -> 73.78 | c.1332-11909C>G | Proximal | Frameshift | Skeletal muscle | [27]       |
| 20   | 18     | 132        | 68.21         | 67.88         | c.2622+1G>A | Distal (3′) | Premature stop codon | Lymphocytes | [10]      |
| Gene Position | Alteration | Identity | 5' Splice Site | Downstream Splice Site | Proximal | Frameshift | Skeletal | Literature |
|---------------|----------------|-----------|----------------|------------------------|----------|------------|----------|------------|
| 32480880-32480945 | tttatattaatcatctctagtGCAAAGGAATTGTTGTTTCCAGTGATGACACA TAATCTGATCGATTTGCTGGCTCCACTCCAGAAGACTGtaaactggttttgaatgta | 66 | 80.03 | 78.22 | None | N/A | ORF preserved | Skeletal muscle | [14] |
| 32461308-32461402 | tatctgtgctttctctaca (a→g)G TATCAGCTCTGGCGTCTGACTTTGACAAATGAGTTAGGTTGTAAAAGGAAGGAACAATGGCGCTCAAGGAGAAGAAGAAAGACGATGCGGTAAAAACAAGGAAGCCATATGTGAATATTGTTACCAATTCAGCATTCCAGAGAGAATAATGGAAATGAAGTGTAAATCTATGCATTACAGAAATATCTACAGACAAA | 95 | 79.15 | 70.75 | c.3432+2036A>G | Proximal | Frameshift | Skeletal muscle | [28] |
| 32461200-32461401 | atctgtgctttctctacaag TATCACTCTGGCGCTCTGACTTTGACAAATGAGTTAGGTTGTAAAAGGAAGGAACAATGGCGCTCAAGGAGAAGAAGAAAGACGATGCGGTAAAAACAAGGAAGCCATATGTGAATATTGTTACCAATTCAGCATTCCAGAGAGAATAATGGAAATGAAGTGTAAATCTATGCATTACAGAAATATCTACAGACAAA (a→g)taagtctgctgtacactctgccatgtttccatctctctctctctctctctctctctctctctctctctctctctctctctctctctc | 202 | 76.71 | 70.14 | c.3432+2240A>G | Proximal | Frameshift | Skeletal muscle | [29] |
| 32452550-32452629 | cttcctcctcctctgtatcagAAAC TGCAATTCCAGATA (G→C) GTCAA ATGTTTGACTGCAATGCAACTTT TATTGCTGATAGCCAGACTGCAACATTT GAGgtatattttttctatctatc | 80 | 86.2 | 77.21 | c.3603+2053G>C | Proximal | Frameshift | Skeletal muscle | [30] |

(Continued)
| PE # | Intron | Sequence | Genomic Co-ordinates (chrX.) | Size (bp) | Acceptor Score | Donor Score | Mutation(s) | Orientation of Mutation to PE and Splice Sites | ORF | RNA Source | Reference |
|------|--------|----------|-----------------------------|----------|----------------|-------------|-------------|-----------------------------------------------|-----|------------|----------|
| 26   | 27     | catatctctatgtgttacagATGT#TGGTGTCTTCTTATTTGATGCTGTTTTTTCATTTACTGATACATTGAGTTGCTTCTAATTTTAGATGACTATCTTCCTCTTACATTACGgt(t>c>a)agtttttaagagatgtg | 32442160-32442278 | 119 | 87.46 | 80.58 -> 90.39 | c.3787-843C>A | Proximal | Frameshift | Lymphocytes or skeletal muscle (unspecified) | [31] |
| 27   | 29     | cattttttttctatcaacagAGCTGAGAGTTGTTTTCTTCTATTTTGAGTCTGTTTTTCAATTACTTGATGGTGTCCTTTGATGAATAGGTCCTAATTTTAATGTAGTCGAATGTATTCATCTTTTTTTTCTTACG | 32412019-32412063 | 45 | 90.17 | 83.09 | c.3613delG | Distal (5’) | Premature stop codon | Lymphocytes | [10] |
| 28   | 34/42  | caaatgctctctcatacatagATGT#TGGCCTTTATAGATGACACAAATAAGCAGATGCTGCTGCCAGCTGGATGATGCTTACTGAGGCCTATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG}
| 32-43 | ttttatatatatatatttattagATTTC | chr4:182041664-182041743 | chr4:182041664-182041743 | 80 | 90.26 | 99.52 | Complex insertion/deletion (see cited work) | Encompassing | Frameshift | Skeletal muscle | [36] |
| 33-43 | tttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
| PE # | Intron | Sequence | Genomic Co-ordinates (chrX.) | Size (bp) | Acceptor Score | Donor Score | Mutation(s) | Orientation of Mutation to PE and Splice Sites | ORF | RNA Source | Reference |
|------|--------|----------|----------------------------|-----------|----------------|-------------|-------------|-----------------------------------------------|-----|------------|-----------|
| 36   | 45     | tttcttcggagtattctagGAGAA | 31965031-31965167 | 137 | 73.52 | 71.16->88.39 | c.6614+3310G>T | Proximal | Frameshift | Skeletal muscle | [27] |
| 37   | 47     | aattatggtaatccccaca(t>g)G | 31879338-31879409 | 72 | 61.61->78.18 | 74.53 | c.6913-4037T>G | Proximal | Premature stop codon | Skeletal muscle | [27] |
| 38   | 48     | aaagctgctctttccgctagGOTCC | 31850637-31850744 | 108 | 81.97 | 70.41 | Complex inversion (see cited work) | Encompassing | Premature stop codon | Cultured myogenic cells | [39] |
| 39   | 48     | gacagtggctcttattagCCAAAT | 31845152-31845276 | 125 | 75.76 | 78.66 | Complex inversion (see cited work) | Encompassing | Frameshift | Cultured myogenic cells | [39] |
| 40   | 49     | atgtatggtaccttttacagGAAGT | 31833559-31833718 | 180 | 93.4 | 87.02 | Complex inversion (see cited work) | Encompassing | Premature stop codon | Cultured myogenic cells | [39] |
| Frame | Nucleotides | Location | Encompassing Frameshift | Complex Inversion (see cited work) | Cultured myogenic cells |
|-------|-------------|----------|-------------------------|-------------------------------------|------------------------|
| 41-49 | atgatgcgtgtcattttgcagTTCAT | 31832500-31832659 | 160 87.27 75.66 | Complex inversion (see cited work) | Encompassing Frameshift Cultured myogenic cells |
| 42-49 | gtaagtaaaacatattacta | 31831990-31832138 | 149 88.15 88.06 | Complex inversion (see cited work) | Encompassing Frameshift Cultured myogenic cells |
| 43-49 | tattgttatcttctttctagAAGTC | 31831041-31831138 | 98 73.62 87.58 | Complex inversion (see cited work) | Encompassing Frameshift Cultured myogenic cells |
| 44-51 | aaatggtcttttggttacagGGTGA | 31752441-31752524 | 84 96.74 76.2 | None N/A ORF preserved Skeletal muscle |  |
| 45-51 | tgcctttatgatattttaagAATTC | 31765009-31765013 | 103 79.68 68.7 | c.7542+8951 7542+8952 ins6091 | Proximal Frameshift Skeletal muscle |

(Continued)
| PE # | Intron # | Sequence | Genomic Co-ordinates (chrX) | Size (bp) | Acceptor Score | Donor Score | Mutation(s) | Orientation of Mutation to PE and Splice Sites | ORF | RNA Source | Reference |
|------|----------|----------|-----------------------------|-----------|----------------|-------------|-------------|-----------------------------------------------|-----|------------|-----------|
| 46   | 53       | ttatatcctgaggaattatatagACTAC TAAAGCACACAGATATTTTGGGAAAT TCCAGGAGAAGAGGAAAGAGAAGAGAAGAA GAGGAGAAGCTGGGCTGGAGgtaagag aatgggggggaa | 31678558-31678630 | 73     | 64.29          | 91.82       | Complex inversion (see cited work) | Encompassing Frameshift | Skeletal muscle | [41]     |
| 47   | 55       | tttacttttttcatttaca(a>g) AATTGAAACACCAAGAAAACAG AATTATGGAAGAGGAAGAAAGa | gtaagag aatgggggggaa | 31609571-31609620 | 50     | 79.61 ->       | 96.18       | c.8217+ 18052A>G | Proximal Frameshift | Skeletal muscle | [33]     |
| 48   | 55       | tttttttttttcttttttagAGTTC TGCTAATGAGGGGACAAAGTTAT ATTAAGAACGTGAAAGTTAATTCCA ACCAATTCTTTTATCGAggtg+g | gtaagtagccagat | 31595572-31595644 | 73     | 90.26          | 78.52 -> 95.74 | c.8217 +32103G>T | Proximal Frameshift | Skeletal muscle | [35]     |
| 49   | 56       | ctcatcttcctaccagAAAAT | gtaatgaggtt | 31497079-31497244 | 166    | 94.32          | 87.36       | c.8391-1419,8391-828del; c.8391-102,8391-76del | Distal (5' and 3') Frameshift | Skeletal muscle | [37]     |
| 50   | 60       | tttttttttttttttgttcagGTTGA GTCCTCAGAAGCAGATCCGACCG GAGATTATTCGAGAATAACCTTG TGAGAAGATAAGGGTGGAGCAAAGAA TTGACAAAGggtg+aagccatcag atcacaat | gtaatgaggtt | 31364155-31364243 | 89     | 88.65          | 76.64 -> 93.87 | c.9085-15519G>T | Proximal Frameshift | Skeletal muscle | [9]      |
| 51   | 62       | ttatatatatatatatgtgcagAGAGCT GATAGCCACACACACTTCAAGA AATTGAGAAGACGTGGAAGCTGTC ATCCAGGCCA(A>G)gtaatata ggaagaggtgt | gtaatgaggtt | 31261663-31261729 | 67     | 96.61          | 75.91 -> 88.06 | c.9225-647A>G | Proximal Frameshift | Skeletal muscle | [9, 42]  |
| SNP  | Type   | [c.9225-285A>G](https://www.ncbi.nlm.nih.gov/snp/9225-285A>G) | [c.9563+1215A>G](https://www.ncbi.nlm.nih.gov/snp/9563+1215A>G) | [c.9564-427T>G](https://www.ncbi.nlm.nih.gov/snp/9564-427T>G) | [c.9807+2714C>T](https://www.ncbi.nlm.nih.gov/snp/9807+2714C>T) |
|------|--------|-------------------------------------------------------------|-------------------------------------------------------------|-------------------------------------------------------------|-------------------------------------------------------------|
| 52   | Proximal Frameshift   | Skeletal muscle [28]; Lymphocytes or skeletal muscle (unspecified) [31] | Lymphocytes [10] |  |  |
| 53   | Unknown Premature stop codon | Unknown | Unknown | Premature stop codon | Lymphocytes [10] |
| 54   | Proximal Premature stop codon | Skeletal muscle |  |  |  |
| 55   | Proximal Frameshift | Skeletal muscle |  |  |  |
| 56   | ORF preserved | Lymphocytes [10] |  |  |  |
| 57   | Premature Frameshift | Lymphocytes or skeletal muscle (unspecified) [31] |  |  |  |
| 58   | Unknown Frameshift | Lymphocytes [10] |  |  |  |

**References:**

[28, 31] 28, 31

**Gene:** DMD

**Mutation:**

- *tgtcgtggttctctctcttgtag* → *TGTTC*
- *tttttctctctccttgtag* → *TTACA*
- *ttttttctttcttttcag* → *TTACA*
- *gtgttttgttttgttttcag* → *TTACA*
- *gtacttatggtcaattgcag* → *TTACA*

**Mutation Type:**

- 52 62: Proximal Frameshift
- 53 63: Unknown Premature stop codon
- 54 65: Proximal Premature stop codon
- 55 65: Proximal Frameshift
- 56 67: ORF preserved
- 57 67: Premature Frameshift
- 58 77: Unknown Frameshift
exons of the gene – though, as with PE04, it is yet to be determined whether these inclusions indicate a functional role for PE07 [17].

**PSEUDOEXONS OF UNKNOWN ORIGIN**

Underlying genomic mutations were not identified for three of the pseudoexons catalogued (PEs 53, 56 and 58). However, as these pseudoexons were exclusively detected in the RNA of specific DMD patient cells, they are believed to be pathogenic and therefore were not classed as arising *sans* mutation.

**PSEUDOEXONS ARISING FROM DELETIONS**

Of the 54 known mutation-obligate DMD PEs, ten arose from genomic deletions: PEs 02, 03, 12, 17, 18, 27, 28, 29, 35, and 49. For some of these cases, PE initiation can easily be explained as a direct result of the deletion event bringing into conjunction tracts of sequence that, when transcribed, present a strong exon signal to the spliceosome. Pseudoexons appearing to fit this description are PEs 17, 28 and 35, though it should be noted that PE28 also has two small insertions (17bp and 8bp) near its junction site. Two additional but less obvious examples can be seen with PEs 29 and 49 – in these cases, the sequence of the pseudoexons and their splice sites are unaltered from normal individuals, but their inclusions in immature transcripts are initiated by deletions of immediately flanking intronic regions, which presumably contain essential splicing silencers.

For the remaining five deletion-initiated PEs, the link between mutation and pseudoexon is less clear. PEs 12 (i3), 18 (i11) and 27 (i29) all arose from frameshifting deletions of one, two and one bases in exons 5, 12 and 27 respectively, and PE02 (i1) and PE03 (i1) (both from the same patient) were purportedly initiated by a deletion of exons 3 to 6. Though a more detailed explanation of these PEs may not be possible at present, they appear to support the general theory that splicing of a given DMD intron is often interdependent on the correct processing of distant elements of the same transcript [46].

**PSEUDOEXONS ARISING FROM Duplications**

Five DMD pseudoexons arose from genomic duplications: PEs 08, 09, 10, 11, and 33. PE08 (i2), which has also been observed as a low-frequency inclusion in normal skeletal muscle RNA, was converted to a pseudoexon by a tandem duplication of exon 2. PE09 (i2), PE10 (i2) and PE11 (i3) were reported in the same proband as a result of an exon 8–11 duplication, and PE33 (i43) arose from an exon 44 duplication. These cases offer further support to the theory of correct DMD splicing occurring through coordination of distant elements. At this point, however, it is not clear whether these PEs are induced specifically by alterations to the canonical exon order, or whether they arise from disruptions to intronic sequences that would normally act as distal pseudoexon silencers.

**PSEUDOEXONS ARISING FROM INVERSIONS**

Eight DMD pseudoexons arose from inversion mutations: PEs 34, 38, 39, 40, 41, 42, 43, and 46. In all these cases, each PE was completely internal to the inverted region. PE34 arose from an inversion internal to intron 44 – i.e. no canonical exons were directly affected. PEs 38 to 43 (i48 and i49) were reported from a single patient with a complex inversion of exons 49 and 50, while PE46 (i53) arose in a patient with a deletion of exons 48–52 and an inversion of exon 53. It is perhaps unsurprising that such dramatic rearrangements of large tracts of transcribed sequence would result in splicing disturbances of some kind, but these cases nevertheless serve to illustrate that, in addition to recognition of canonical exons, the silencing of pseudoexons is an equally essential component of spliceosome function, and one that is likely to be achieved through orientation-dependent sequence motifs in the intron.

**PSEUDOEXONS ARISING FROM INSERTIONS**

Two DMD pseudoexons arose from insertion mutations, PEs 32 and 45. PE32 was created by an insertion into intron 43 of two large tracts (88.0kb and 2.6kb) of intragenic sequence from chromosome 4, the PE itself originating within the larger of these two tracts, while PE45 was created by a 6096bp LINE-1 retrotransposon with a potential donor site at its 5′ end inserting immediately 3′ of a latent acceptor site in DMD intron 51.

**PSEUDOEXONS ARISING FROM SINGLE BASE-PAIR SUBSTITUTIONS**

Single base-pair substitutions were the most commonly observed cause of DMD PEs, accounting for
Fig. 1. Locations of pseudoexon-initiating single-nucleotide variations in the DMD gene, relative to acceptor and donor splice site consensus sequences. Numbers above each nucleotide indicate the exemplar pseudoexons. Lower-case letters indicate intron sequence, upper-case letters indicate exon sequence. Dash-line boxes highlight the essential “ag” and “gt” of the acceptor and donor site motifs respectively.

26 of the 58 catalogued, through 24 unique mutations. In most of these cases, the etiology of the PE appears to stem from the creation or enhancement of a latent mid-intron splice site – of the 24 unique mutations, 7 created new acceptor splice sites (PEs 01, 05/06, 15, 22, 30, 37 and 47) and 15 created new donor splice sites (PEs 13, 14, 16, 19, 23/24, 26, 31, 36, 48, 50, 51, 52, 54, 55 and 57). All 22 of these acceptor-motif and donor-motif mutations greatly enhanced the Shapiro-Senapathy splice score of the mutated site, and in every case the new nucleotide was the most common consensus base for that position in the splice site (Fig. 1). While a possible exception to this rule was noted at the acceptor site of PE14 (c.650-39575A>C), this mutation was found to be a common SNP (rs113593006, dbSNP build 151 – see ref. 44) that only marginally decreased the Shapiro-Senapathy score of the acceptor site (from 79.91 to 78.04). I therefore judged that this SNP was likely to be incidental to the pathology of this pseudoexon and did not constitute a true counterexample to the prevailing pattern of splice site enhancement.

Only two PEs arose from SNVs outside of the PE consensus splice sites, PE20 and PE25. PE20 (i18) arose from a G-to-A substitution at the first base of intron 20, suggesting that the correct splicing of these two introns may be interdependent. In this way, PE20 is qualitatively similar to PEs 12 and 18, which also arose from small mutations distal to the pseudoexon, although the mutation that initiated PE20 did not directly alter the DMD coding sequence. PE25 (i26) is a unique case that arose from a G-to-C substitution internal to the PE that altered the predicted binding of splicing enhancer SRp55 [30].

PSEUDOEXONS AND RECURSIVE SPLICING

Multi-step or recursive splicing was first described in Drosophila in 2005 [47] and has more recently been discovered to be prevalent in the genes of the human transcriptome [48, 49], including DMD [50, 51]. While conventionally spliced introns are removed with a single splicing event, recursively spliced introns are excised from their maturing transcripts in two or more segments, via intronic acceptor splice sites called ‘ratchet points’. Sibley et al. [48] have also reported that recursive splicing in vertebrates is facilitated by recognition of evolutionarily conserved donor-like splice sites downstream of acceptor-like ratchet points.

Georgomanolis et al. [49] have postulated that some of the low-frequency pseudoexons observed in the transcripts of normal cells may be a natural byproduct of the spliceosome incorrectly recognizing exon-like intronic ratchet points. I suggest that this hypothesis can reasonably be extended to include mutation-induced PEs – i.e. mutations that enhance the exon-like characteristics of intronic ratchet points may thereby convert them into pathogenic PEs. Evidence supporting this hypothesis has already been described by Bouge et al. [14], who noted the alignment of six pseudoexon splice sites with six of the DMD intron ratchet points predicted by Gazzoli et al. [51]. Seeking to expand upon these observations, I cross-referenced the splice sites of all eligible PEs in Table 1 with all of the intronic ratchet points predicted by Gazzoli et al. This analysis excluded the splice sites of the DMD inversion PEs (34, 38, 39, 40, 41, 42, 43 and 46), the chromosome 4 insertion PE (32), and the de novo donor site for PE45, as these sites could not be sensibly compared to any part of the DMD reference sequence. Splice sites shared by multiple PEs (acceptor sites for PEs 5 and 6, 15 and 16, and donor sites for PEs 9 and 10, 11 and 12, and 15 and 16) were included but were counted only once each to avoid bias. Using these criteria, including the matches noted by Bouge et al. I confirmed 12 Gazzoli matches out of 47 unique acceptor sites and 14 Gazzoli matches out of 44 unique donor sites (Table 2).
Several interesting features were apparent in this set of Gazzoli-matched splice sites. Firstly, most of the matched PEs matched at both their acceptor and donor splice sites. Only PEs 1, 11/12 and 49 matched at their donor splice sites alone, and only PE56 matched at its acceptor site alone. Secondly, a clear bias was evident in the mutation categories of the matched PEs, as the majority of the Gazzoli-matched sites were from PEs induced either without mutations or by mutations distal to the PE and its splice motifs, PEs 1 and 26 being the only exceptions to this rule. Lastly, of the 15 PEs where inducing distal mutations were identified, 11 arose exclusively from mutations that were 3’ to the PE (PEs 2, 3, 4, 7, 8, 9, 10, 11, 12, 18 and 20). Only PEs 27 and 29 were induced exclusively by 5’ mutations, while PEs 33 and 49 were each induced by flanking mutations.

**PSEUDOEXONS AND RECURSIVE SPlicing REGULATION**

Canonical splicing of a donor-acceptor pair is often dependent on distal regulatory elements, including
Fig. 2. Suggested model of the two most common modes of pseudoexon initiation observed in the DMD gene. (A) Proximal mutations at non-RS sites. (i) In the absence of mutation, a putative pseudoexon presents a weak exon-like profile to the spliceosome and is predominantly excluded from mature transcripts. (ii) The presence of a mutation, usually a splice-site-creating SNV, increases the exon-like profile of the putative pseudoexon, resulting in its inclusion in a much higher proportion of transcripts. (B) Mutations 3' distal to RS sites. (i) In the absence of mutation, the exon n donor site and RS-acceptor site are used to excise the 5' segment of an intron. Silencing elements distal (usually 3') to the RS-exon prevent spliceosome recognition of its donor-like motif, and the RS-exon is subsequently removed from the transcript along with the rest of the 3' intron segment. (ii) When mutations to the distal silencing elements impair their function, the intron segment 5' to the RS-exon is spliced as normal, but the RS-exon donor-like site escapes silencing and is more readily recognised by the spliceosome, leading to a much higher frequency of inclusion of the RS-exon in the mature transcript population.

but not limited to other canonical splice sites [51, 52, 53]. Mutations that alter or destroy these distal elements can impede exon definition and decrease the frequency of inclusion of the affected exons in the mature transcript [42, 43]. It may be that spliceosome recognition of recursive splice sites (which necessarily exhibit a strong exon-like profile in their local motifs) is regulated by a similar system of mostly 3' distal elements, but a system that acts to silence rather than promote the inclusion of its targets. Mutations that impaired these distal silencing elements might thereby permit an increase in the erroneous processing of recursive splice sites, converting them to PEs via a distinctly different pathway than PEs created by proximal mutation (Fig. 2). If valid, this model would explain the high coincidence of Gazzoli predicted recursive splice sites with the splice sites of PEs induced by distal mutations. However, while the mutations collated in this report may offer broad clues as to the locations of some such suppressive distal elements in the DMD gene, any further analysis of their common features awaits the assembly and analysis of a much larger dataset of PEs and recursive splice sites – one that will have to encompass multiple other genes besides DMD.

CONCLUSIONS

The 58 DMD pseudoexons collated from published reports exhibit great diversity in their sizes, locations, and pathologies. Surprisingly, PEs arising either from no mutation, or from mutations distal to the pseudoexon and its splice sites, exhibited a high coincidence with predicted recursive splice sites in the DMD introns, suggesting that some such pseudoexons may arise from disruptions to recursive splicing regulation. This finding may represent an important new insight into the etiology of pseudoexons in DMD specifically and human disease genes generally.
ACKNOWLEDGMENTS

I would like to thank the Enid and Arthur Home Scholarship and Murdoch University for the provision of funding and resources; my reviewers, for their thoroughness and insightful suggestions; and my colleagues at the Centre for Molecular Medicine and Innovative Therapeutics, especially professors Sue Fletcher and Steve Wilton, for their advice and constructive criticism.

CONFLICT OF INTEREST

The author has no conflict of interest to declare.

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: https://dx.doi.org/10.3233/JND-190431.

REFERENCES

[1] Nudel U. Alternative Promoters: Duchenne Muscular Dystrophy (DMD) Gene. In eLS, (Ed.), 2005.
[2] Feener CA, Koenig M, Kunkel LM. Alternative splicing of human dystrophin mRNA generates isoforms at the carboxy terminus. Nature. 1989;338(6215):509-11.
[3] Doorenweerd N, Mahfouz A, van Putten M, Kaliyaperumal R, t’Hoen PAC, Hendriksen JGM, et al. Timing and localization of human dystrophin isoform expression provide insights into the cognitive phenotype of Duchenne muscular dystrophy. Sci Rep. 2017;7(12575).
[4] Wahl MC, Will CL, Lührmann R. The Spliceosome: Design Principles of a Dynamic RNP Machine. Cell. 2009;136:701-18.
[5] Berget SM. Exon recognition in vertebrate splicing. J Biol Chem. 1995;270(6):2411-4.
[6] De Conti LM, Baralle M, Buratti E. Exon and intron definition in pre-mRNA splicing. Wiley Interdiscip Rev RNA. 2013;4(1):49-60.
[7] Anna A, Gos M. Splicing mutations in human genetic disorders: examples, detection, and confirmation. J Appl Genet. 2018;59(3):253-68.
[8] Hong X, Scofield DG, Lynch M. Intron size, abundance, and distribution within untranslated regions of genes. Mol Biol Evol. 2006;23(12):2392-404.
[9] Beroud C, Carrie A, Beldjord C, Deburgrave N, Llesson S, Carelle N, et al. Dystrophinopathy caused by mid-intronic substitutions activating cryptic exons in the DMD gene. Neuromuscular disorders: NMD. 2004;14(1):10-8.
[10] Zhang Z, Habara Y, Nishiyama A, Oyazato Y, Yagi M, Takeshima Y, et al. Identification of seven novel cryptic exons embedded in the dystrophin gene and characterization of 14 cryptic dystrophin exons. Journal of human genetics. 2007;52(7):907-17.
[11] Roberts RG, Bentley DR, Bobrow M. Infidelity in the structure of ectopic transcripts: a novel exon in lymphocyte dystrophin transcripts. Human mutation. 1993;2(4):293-9.
[12] Suminaga R, Takeshima Y, Adachi K, Yagi M, Nakamura H, Matsuo M. A novel cryptic exon in intron 3 of the dystrophin gene was incorporated into dystrophin mRNA with a single nucleotide deletion in exon 5. Journal of human genetics. 2002;47(4):196-201.
[13] Tran VK, Zhang Z, Yagi M, Nishiyama A, Habara Y, Takeshima Y, et al. A novel cryptic exon identified in the 3' region of intron 2 of the human dystrophin gene. Journal of human genetics. 2005;50(8):425-33.
[14] Bouge AL, Murauer E, Beyne E, Miro J, Varilh J, Taulan M, et al. Targeted RNA-Seq profiling of splicing pattern in the DMD gene: exons are mostly constitutively spliced in human skeletal muscle. Sci Rep. 2017;7:39094.
[15] Yagi M, Takeshima Y, Wada H, Nakamura H, Matsuo M. Two alternative exons can result from activation of the cryptic splice acceptor site deep within intron 2 of the dystrophin gene in a patient with as yet asymptomatic dystrophinopathy. Human genetics. 2003;112(2):164-70.
[16] Dwi Pramono ZA, Takeshima Y, Surono A, Ishida T, Matsuo M. A novel cryptic exon in intron 2 of the human dystrophin gene evolved from an intron by acquiring consensus sequences for splicing at different stages of anthropoid evolution. Biochemical and biophysical research communications. 2000;267(1):321-8.
[17] Gualandi F, Rimessi P, Trabancelli C, Spitali P, Neri M, Patarrello T, et al. Intronic breakpoint definition and transcription analysis in DMD/BMD patients with deletion/duplication at the 5′ mutation hot spot of the DMD gene. Gene. 2006;370:26-33.
[18] Ishibashi K, Takeshima Y, Yagi M, Nishiyama A, Matsuo M. Novel cryptic exons identified in introns 2 and 3 of the human dystrophin gene with duplication of exons 8-11. The Kobe journal of medical sciences. 2006;52(3-4):61-75.
[19] Sedláčková J, Vondráček P, Hermanová M, Zámečník J, Hrubá Z, Haberlová J, et al. Point mutations in Czech DMD/BMD patients and their phenotypic outcome. Neuromuscular Disorders. 2009;19(11):749-53.
[20] Zaum AK, Stuve B, Gehrig A, Kolbel H, Schara U, Kress W, et al. Deep intronic variants introduce DMD pseudoexon in patient with muscular dystrophy. Neuromuscular disorders: NMD. 2017;27(7):631-4.
[21] Tuffery-Giraud S, Saquet C, Thorel D, Disset A, Rivier F, Carel N, et al. Dystrophinopathy caused by mid-intronic nucleotide deletion in exon 5. Journal of human genetics. 2008;53(12):1254-60.
[22] Ferlini A, Muntoni F. The 5′ region of intron 11 of the dystrophin gene contains target sequences for mobile elements and three overlapping ORFs. Biochemical and biophysical research communications. 1998;242(2):401-6.
[23] Ferlini A, Galie N, Merlini L, Sewry C, Branzi A, Muntoni F. A novel Alu-like element rearranged in the dystrophin gene causes a splicing mutation in a family with X-linked dilated cardiomyopathy. American journal of human genetics. 1998;63(2):436-46.
[24] Nasim MT, Chernova TK, Chowdhury HM, Yue BG, Eperon IC. HnRNP G and Tra2beta: opposite effects on splicing matched by antagonism in RNA binding. Human molecular genetics. 2003;12(11):1337-48.
[25] Rimessi P, Fabris M, Bovolenta M, Bassi E, Filzarano S, Gualandi F, et al. Antisense modulation of both exonic and intronic splicing motifs induces skipping of a DMD pseudo-exon responsible for x-linked dilated cardiomyopathy. Human gene therapy. 2010;21(9):1137-46.
[26] Maluenga RG, Takaoka Y, Yagi M, Awano H, Lee T, Dwaningsih EK, et al. Categorization of 77 dystrophin exons into...
5 groups by a decision tree using indexes of splicing regulatory factors as decision markers. BMC genetics. 2012;13:23. [27] Gurvich OL, Tuohy TM, Howard MT, Finkel RS, Medne L, Anderson CB, et al. DMD pseudoexon mutations: splicing efficiency, phenotype, and potential therapy. Annals of neurology. 2008;63(1):81-9. [28] Tuffery-Giraud S, Saquet C, Chambert S, Claustres M. Pseudoexon activation in the DMD gene as a novel mechanism for Becker muscular dystrophy. Human mutation. 2003;21(6):608-14. [29] Ikezawa M, Minami N, Takahashi M, Goto Y, Miike T, Nonaka I. Dystrophin gene analysis on 130 patients with Duchenne muscular dystrophy with a special reference to muscle mRNA analysis. Brain & development. 1998;20(3):165-8. [30] Trabetsi M, Beugnet C, Deburgrave N, Commerc V, Orhant L, Leturcq F, et al. When a mid-intronic variation of DMD gene creates an ESE site. Neuromuscular Disord. 2014;24(12):1111-7. [31] Takeshima Y, Yagi M, Okizuka Y, Awa H, Zhang Z, Yamauchi Y, et al. Mutation spectrum of the dystrophin gene in 442 Duchenne/Becker muscular dystrophy cases from one Japanese referral center. Journal of human genetics. 2010;55(6):379-88. [32] Ishmukhametova A, Van Kien PK, Mechlin D, Thorel D, Vincent MC, Rivier F, et al. Comprehensive oligonucleotide array-comparative genomic hybridization analysis: new insights into the molecular pathology of the DMD gene. European Journal of Human Genetics. 2012;20(10):1096-100. [33] Bovolenta M, Neri M, Fini S, Fabris M, Trabaneli C, Venturoli A, et al. A novel custom high density-comparative genomic hybridization array detects common rearrangements as well as deep intronic mutations in dystrophinopathies. BMC genetics. 2008;9. [34] Gonorazky H, Liang M, Cummings B, Lek M, Micallef A, Sibley CR, Emmett W, Blazquez L, Faro A, Haberman N, Briese M, et al. Recursive splicing in long vertebrate genes. Nature. 2015;521(7552):371-5. [35] Georgomaioris T, Sofoadis K, Papantonis A. Cutting a Long Intron Short: Recursive Splicing and Its Implications. Front Physiol. 2016;7:598. [36] Suzuki H, Aoki Y, Kameyama T, Saito T, Masuda S, Tanihata J, et al. Endogenous Multiple Exon Skipping and Back-Splicing at the DMD Mutation Hotspot. International Journal of Molecular Sciences. 2016;17(10). [37] Burnette JM, Miyamoto-Sato E, Schaub MA, Conklin J, Lopez AJ. Subdivision of large introns in Drosophila by recursive splicing at nonexonic elements. Genetics. 2005;170(2):661-74. [38] Sibley CR, Emmett W, Blazquez L, Faro A, Haberman N, Briese M, et al. Recursive splicing in long vertebrate genes. Nature. 2015;521(7552):371-5. [39] Geurtsquant R, Sofoadis K, Papantonis A. Cutting a Long Intron Short: Recursive Splicing and Its Implications. Front Physiol. 2016;7:598. [40] Goncalves A, Oliveira J, Coelho T, Taipa R, Melo-Pires M, Sousa M, et al. Exonization of an Intronic LINE-1 Element Causing Becker Muscular Dystrophy as a Novel Mutational Mechanism in Dystrophin Gene. Genes. 2017;8(10).