Splicing Behaviour And Exotic Mutations In The *DMD* Gene

Niall Keegan, BSc (Hon) Genetics

This thesis is presented for the degree of Master of Philosophy of Murdoch University, 2016

I declare that this thesis is my own account of my research and contains as its main content work which has not previously been submitted for a degree at any tertiary education institution.

Niall Keegan

.....

Abstract

DMD is the largest gene in the human genome, spanning over 2.2Mb of the X chromosome, and more than 99% of the gene content is intronic sequence. *DMD* encodes dystrophin, a crucial protein for protecting muscle fibres from mechanical damage. Mutations to *DMD* can cause any one of a family of diseases, known as the dystrophinopathies. The most severe of these is Duchenne muscular dystrophy, a progressive and global muscle wasting disease that is fatal to affected males in early life. Therapies for dystrophinopathies have progressed substantially in recent years, but at present there is no cure.

The projects comprising my MPhil research aim to improve our understanding of splicing and mutation in *DMD* transcripts. *DMD* splicing is necessarily complex due to the size of the gene and its multiple spliceoforms. As such, the full effects of a given mutation within the gene can be unpredictable.

Because the first step of describing a mutation should be elucidation of the sequence, a new method (Fractal PCR) has been devised for efficiently determining the intronic breakpoints of whole-exon deletions in *DMD.* Existing research into *DMD* and *NF1* pseudoexons was used to inform PCR designs, and this strategy successfully discovered rare alternative transcripts of these two genes in normal human RNA.

For the bioinformatics component of this project, results were compiled for hundreds of putative exon-skipping antisense oligomers (AOs) and a search was conducted for patterns in the splice factor (SF) motifs targeted or avoided by the most effective of these molecules. The intent of this work was to generate a predictive model for optimal AO design. While the power of this model was equivocal in some regards, the characteristics of the SF motifs identified as targets unexpectedly point to a clear link between *DMD* transcript splicing and myotonic dystrophy.

Table of Contents

Acknowledgements

I would like to thank:

The administrators and funders of the WANRI & Enid & Arthur Home Scholarship, without whose financial support none of this work would have been possible;

My supervisors, Prof. Sue Fletcher, Prof. Steve Wilton, Prof. Matt Bellgard, Dr. Roberto Barrero and Dr. Mark Davis, for their support and guidance over the last two years;

My co-workers and fellow students, Kristin West, Kane Greer, Loren Price, Ianthe Pitout, May Aung-Htut, Lucy Barrett, Russell Johnsen, and Abbie Adams, for answering my questions and selflessly volunteering their time, materials and expertise whenever needed;

My Dad, for his love, encouragement and philosophical advice over the course of my academic career;

And my Mum, who did not live to see me achieve this but never doubted that I would.

Chapter 1

General Introduction

1.1 Characteristics of the Dystrophinopathies

Dystrophinopathy is a human neuromuscular disease class with three members: X-Linked Dilated Cardiomyopathy (XLDC), Becker Muscular Dystrophy (BMD) and Duchenne Muscular Dystrophy (DMD) (Darras *et al.* 2000). Each disease stems from a different type of dysfunction in the *DMD* gene, which is carried on the X chromosome at locus Xp21. Of the dystrophinopathies, DMD is the most severe, arising as it does from any *DMD* mutation that causes complete or near-complete loss of functional dystrophin protein (Hoffman *et al.* 1987). While treatments exist, all dystrophinopathies are currently incurable.

DMD affects approximately 3 in 10000 male live births (Emery 1991, *in* Emery 2002), and is both the most common and the most severe of the muscular dystrophies. The primary symptom is a progressive deterioration of all muscle tissues in the body, save for the extra ocular muscles (Khurana *et al.* 1995). Mild to moderate intellectual impairment is also seen in some cases, and there is evidence of an inverse relationship between mental effect and age of physical onset (Desguerre *et al.* 2009). Intellectual impairment is often exacerbated in cases where the causative lesion is a deletion that extends into nearby genes such as NROB1 (Barbaro *et al.* 2012) and GKD (Lin *et al.* 2006).

DMD symptoms usually present in early childhood, and by the age of twelve patients are typically unable to walk and are restricted to wheelchairs. Advances in therapy have extended patient lifespans in recent decades, but even with optimal treatment plans few survive beyond their late twenties (Passmano *et al.* 2012). The ultimate cause of mortality is most often some combination of heart failure and pneumonia (Emery 2002), the latter resulting from respiratory tract infections compounded by the patient's inability to cough as their respiratory muscles weaken.

BMD patients present with qualitatively similar symptoms to DMD patients, but with varyingly lesser degrees of severity, sometimes so much so that they are not diagnosed until late in life (Ferreiro *et al.* 2009). In these cases the *DMD* gene still bears a harmful mutation, but the nature of the mutation is such that a dystrophin protein is still produced, at a clinically significant but diminished

quantity and/or quality (Hoffman *et al.* 1988). The severity of an individual BMD case is dependent largely upon the particular mutation that that patient carries (Beggs *et al.* 1991). Cases described in the literature range from being phenotypically very close to DMD (van den Bergen *et al.* 2014) to virtually asymptomatic (Ferreiro *et al.* 2009).

XLDC patients exhibit a myocardial deterioration similar to that of a DMD or severe BMD sufferer, but with normal to near-normal skeletal musculature. Symptoms are progressive, and analogous to those of other cardiomyopathies, and include arrhythmia, high serum creatine kinase levels and rapidly progressive heart failure (Berko and Swift 1987, *in* Nakamura 2015). XLDC sufferers have a highly variable mortality that appears to be dependent on the nature of their mutation (Oldfors *et al.* 1994; Feng *et al.* 2002), though a minimal skeletal muscle involvement means their mobility and quality of life may be substantially better than those of other dystrophinopathy sufferers. However, this same lack of skeletal involvement can make the symptoms of XLDC initially less obvious than those of the other dystrophinopathies, and the prognosis for patients becomes progressively poorer the longer they go without an accurate diagnosis (Nakamura 2015).

While DMD, BMD and XLDC are useful labels, and convenient for the purposes of diagnosis, they describe only a fraction of the variety in symptoms that can arise from *DMD* mutations, and the boundaries between one category and another are not clearly drawn. An individual patient may present with symptoms that defy easy classification. For example, the mental retardation seen in some DMD and BMD patients seems to lack a clear correlation to the severity of their muscle phenotype (Nardes *et al.* 2012). This diversity has necessitated the rise in mutation-specific therapies for these diseases (see section 1.3), and makes essential the investigation of dystrophinopathies at a detailed genetic level, and the impact that different types of *DMD* mutation can have on the transcription, splicing and translation of the gene.

1.2 Genetic Features of the DMD *Gene*

DMD is the largest gene in the human genome, spanning over 2.48Mb, more than 99% of which is intronic sequence (NCBI, 2015). The entire coding

sequence totals 10,500 bases across 79 exons. (See Figs. 1.1 and 1.2). Nine unique alternative promoters, alternative splicing, and an alternative polyadenylation site (Feener *et al.* 1989) produce at least 19 isoforms that are expressed in all muscle and various other cell types throughout the body, including myoblasts, lymphocytes and retinal cells (Leiden Muscular Dystrophy Pages 2006); however, it is the full-length muscle isoform *Dp427,* utilising all 79 standard exons, that is crucial to the dystrophinopathies (Ahn and Kunkel 1993), and it is this isoform that shall be referred to henceforth unless otherwise specified.

The function of the dystrophin protein has been compared to that of a shock absorber; in fact, this is quite a literal description. The N-terminus of dystrophin binds to cytoskeletal actin, while the C-terminus binds to proteins embedded in the sarcolemma (Rybakova *et al.* 2000). The rod domain that forms much of the centre of the protein acts primarily as a mechanical linkage, but has additional roles in actin binding (Rybakova *et al.* 1996; Amann *et al.* 1998) and fine regulation of blood flow to the muscles via nNOS localisation (Lai *et al.* 2009). The sarcolemmal proteins and other proteins that co-localise with dystrophin form the dystrophin-associated protein complex (DAPC), giving the muscle fibres stability and flexibility and minimising the damage they incur with each contraction (Rafael *et al.* 1996; Straub and Campbell 1997). The DAPC also contributes to the control of sarcolemmal membrane permeability, though precisely how it does so is a point of conjecture at present (Allen and Whitehead 2011). If the dystrophin protein is absent, incomplete or otherwise functionless, the muscle fibres are far more vulnerable to mechanical damage. Consequently, their degradation is greatly accelerated (Petrof *et al.* 1993). This damage disrupts calcium ion homeostasis, and triggers the inflammation response and an increase in oxidative stress. As these conditions persist, they mutually reinforce and exaggerate each other (Shin *et al.* 2013), driving the muscle structure to degrade even more rapidly. In the very early years of life, a DMD patient's natural regenerative abilities will somewhat compensate for this degeneration, but the accelerated turnover of cells eventually overwhelms regenerative capacity, leading to fibrosis, a substitution of muscle with fatty tissue and the onset of the aforementioned symptoms (Zacharias and Anderson 1991 *in* Abdel-Salam *et al.* 2009). A comparison of dystrophic and normal

muscle sections (following biopsy and staining) also reveals a predominance and increase in size of type 1 (slow twitch) muscle fibres in dystrophic muscle as the disease progresses (Cavalcanti *et al.* 2011).

A broad range of possible *DMD* mutations and mutation types can produce an identical DMD phenotype (Aartsma-Rus *et al.* 2006). As a rule, any genetic change that disrupts the open reading frame of the gene will lead to a functionless, prematurely truncated polypeptide, or no protein at all. In some cases, however, a mutation occurs that only partially impairs dystrophin function, leading to a BMD phenotype. These mutations can be just as varied in form but, unlike DMD mutations, they typically preserve the reading frame of the mature mRNA (Monaco *et al.* 1988). In discussions of dystrophinopathy genetics, this distinction is referred to as the “reading frame rule”. Though there are numerous exceptions (Kesari *et al.* 2008), these mostly arise when the DNA mutation fails to accurately predict splicing of the transcript.

Individual BMD phenotypes are highly diverse, and the array of resultant symptoms and their severities is dependent on the nature and location of the causative defect (Hamed *et al.* 2005; Anthony *et al.* 2011; van den Bergen *et al.* 2014). BMD mutations have some combination of two effects: reducing the *quantity* of dystrophin produced, and reducing the *quality* of dystrophin produced. The mutations can best be classified as a combination of their nature (e.g. deletion, duplication, inversion, insertion or point mutation), magnitude and location. Of these three factors, location effects are perhaps the best understood. In-frame (BMD) mutations are better tolerated when they occur in the central rod region of the protein (Matsumura *et al.* 1993), and more so if they also preserve the secondary and tertiary stability of the translated protein (Ruszczak *et al.* 2009). To extend the previous metaphor further, such a mutation is equivalent to a shock absorber that has a shortened or distorted spring, but is still securely fixed to the wheel and chassis at either end. Correspondingly, mutations carry a much poorer prognosis when they disrupt the anchoring N-terminal (Prior *et al.* 1993) and C-terminal (Bies *et al.* 1992) domains, and often result in DMD regardless of reading frame preservation, though there are limited exceptions (Yagi *et al.* 2003).

Mutations that impact splicing and exon selection can be particularly cryptic in their effects. SNPs that occur in the immediate vicinity of splice sites can cause splice site slippage, whole exon skipping, or some combination of the two (Takeshima *et al.* 2010; Juan-Mateu *et al.* 2013), but at present there is no way of consistently predicting what the effects will be, based on the sequence data alone. Mutations that induce skipping of an in-frame exon in the *DMD* central rod domain tend to produce a BMD phenotype, frequently a very mild one.

This figure has been omitted from the online version of this thesis so as not to infringe the copyright of its original creator(s).

Figure 1.1: Exons of the DMD gene, with exon sizes and reading frame at the exon-exon junctions. *Adapted from an image created by A. Adams, Perron Institute, Perth, Australia.*

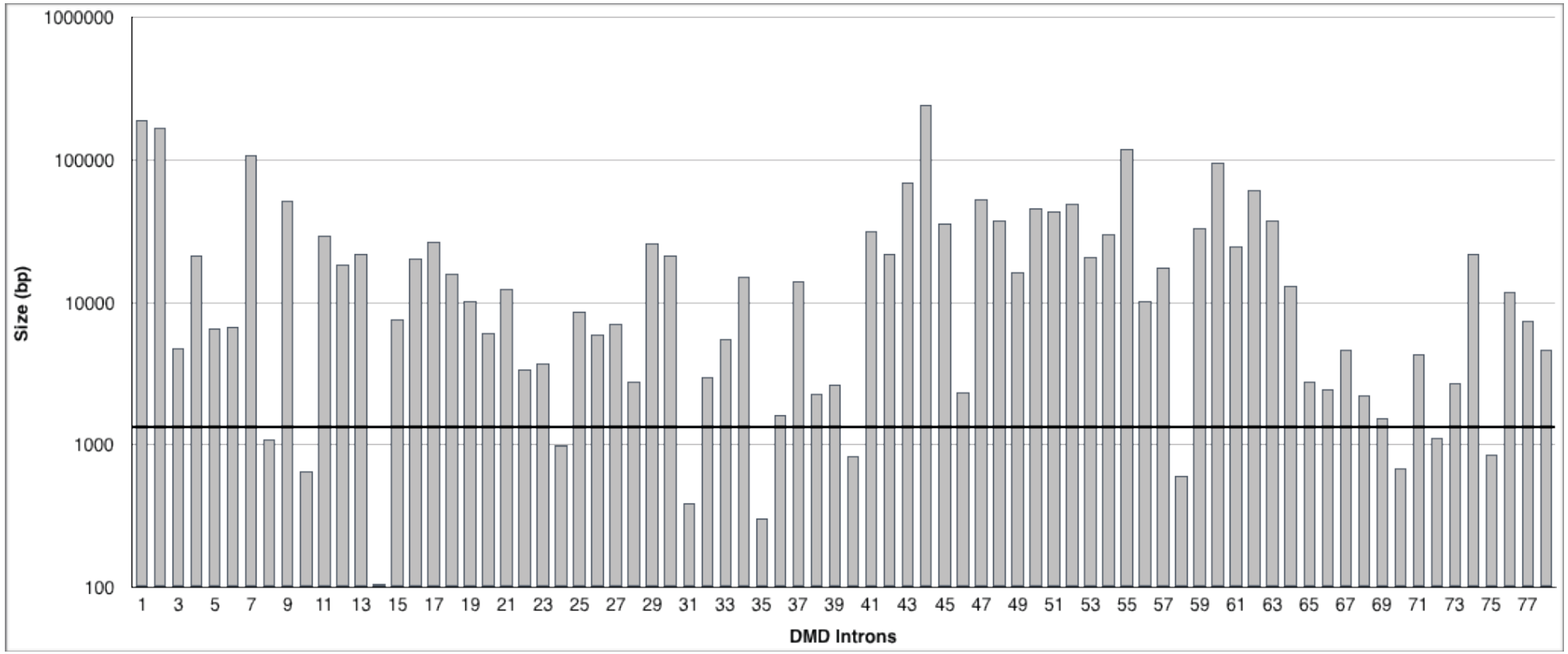


Figure 1.2: Introns of the DMD gene. Sizes are shown on a logarithmic scale. For comparison, the human median intron size (1334bp - Hong *et al.* 2006) is indicated with an overlaid black line

1.3 *Antisense-Oligonucleotide-Mediated Exon Skipping Therapy For Dystrophinopathies*

Exon skipping therapy is designed to exploit the natural capacity of the human spliceosome to omit from the mature transcript (i.e. “skip”) exons that are alternatively spliced, or in some cases that have a mutated splice site (Juan-Mateau *et al.* 2013 and Takeshima *et al.* 2010).(See Fig. 1.3).

Shimitzu *et al.* (1988, in Klein *et al.* 1992) noted that some DMD patients possessed a very small number of revertant, dystrophin-positive muscle fibres that gradually increased in frequency with the patient’s age, though they never reached clinically significant levels (reviewed in Wilton and Fletcher 2010). Revertant fibres were initially thought to be the result of secondary somatic mutations that restored the *DMD* reading frame in some cells (Klein *et al.* 1992), but they later proved to have identical DNA sequence to the surrounding tissue (Lu *et al.* 2000). Revertant fibres spontaneously splice out the mutated exon during mRNA processing, together with flanking exons in some cases, restoring the reading frame and producing a truncated but partly functional dystrophin protein (Sherratt *et al.* 1993). The reason for differential splicing being localised to specific fibres is not yet known, but the current hypothesis is epigenetic inheritance from a single, altered myogenic cell (Lu *et al.* 2000; Yokota *et al.* 2006).

The discovery and characterisation of revertant fibres proved that it was possible for the cells of *DMD* patients to produce functional dystrophin through natural exon skipping. This led some researchers to investigate whether exon skipping could be artificially induced at therapeutic levels. Wilton *et al.* (1999) hypothesised that targeting antisense oligomers (AOs) to the splice sites of mutated exons would mimic the effects of a disruptive splice site mutation and cause that exon to be skipped, thereby producing a BMD-like phenotype. Evaluation of an AO targeted to *DMD* exon 23 in *mdx* mice provided the first *in vivo* proof-of-concept (Mann *et al.* 2001), and since then there has been substantial work done to both expand the library of functional AOs and optimise their design (Adams *et al.* 2007) and chemistry (Heemskerk *et al.* 2009; Mitrpant *et al.* 2009; Wu *et al.* 2012).

The greatest strength of exon skipping as a therapy is its mutation specificity, but this is also a major weakness. Each new AO is effectively a new drug, and must undergo rounds of testing and verification before proceeding to clinical trials. Further complicating this problem is the inconsistent response of different exons to AO intervention. While some exons skip easily, others have proved much more tenacious in their resistance, and the reasons for this are not yet fully understood (Wilton *et al.* 2007). Skipping of resistant exons can sometimes be achieved via the use of AO cocktails, targeting multiple sites within the same exon (Aartsma-Rus *et al.* 2006; Adams *et al.* 2007), but this is only a partial solution.

Our capacity to discover and define *DMD* mutations on a patient-by-patient basis has improved dramatically over the past few decades, and no doubt will continue to be refined even further. As this occurs, the development of a full and detailed understanding of the prognosis entailed by any given mutation will become more vital than ever. The benefits of building this body of knowledge will be twofold: researchers and physicians will be better equipped to design and prescribe treatments tailored to individual patients; and the patients themselves will have a clearer concept of how they can expect their disease to progress, which will better equip them and their families to make realistic long-term plans for their future.

Figure 1.3: Antisense Oligonucleotide (AO or AON) mediated exon skipping in the *DMD* gene transcript. (A) The exon 45-55 section of the full length *DMD* transcript. (B) A patient has a mutation that causes loss of exon 50 in their mRNA. This exon loss causes a frame shift, creating a premature stop codon and preventing the translation of functional protein. (C) An exon-skipping AO (in this case, the drug Eteplirsen) inhibits the exon recognition of exon 51, causing it to be excluded from the mature transcript. (D) The reading frame of the patient transcript is restored, and a shorter but functional isoform of dystrophin can be translated. *(Image credit: Prof. Sue Fletcher, Centre for Comparative Genomics, Murdoch University, Australia)*

1.4 The Complex Behaviour of DMD *Mutations*

While the pathology of some *DMD* mutations is relatively straightforward, many are much more complicated in their aetiology. The following section will outline the current understanding of the types of *DMD* mutations that cause dystrophinopathies, and how well the nature of a given mutation can explain the resultant phenotype.

1.4.1 Large Exonic Deletions

Deletions of one or more exons of the *DMD* gene are responsible for approximately 60% of DMD and BMD cases (Sherrat *et al.* 1993). Larger deletions can often encompass multiple exons, but are highly variable in the locations of their (usually intronic) breakpoints. Intriguingly, even deletions that span identical sets of exons can produce dramatically different phenotypes. In their 2003 review of the topic, Muntoni *et al.* noted:

> "In our experience of 13 patients with an exon 44 deletion, eight had DMD, four had BMD, and one had an intermediate phenotype between DMD and BMD. Of nine patients with an exon 45 deletion, seven developed DMD and two had BMD."

While some of this variation may be due to the deletion breakpoints directly disrupting the expression of other dystrophin isoforms, or sequence variation in other genes, it is also an indication that regulatory elements within the introns could be differentially disturbed from one patient to the next. However, it is clear that alterations to the *DMD* coding sequence alone are not sufficient to predict the phenotype of the patient. Ideally, the genomic breakpoints of dystrophinopathy patients with large deletions and unexpected phenotypes would always be sequenced as a matter of course. In reality, however, such detailed sequencing is rarely practical, and even if it were, making use of the sequence data generated would require a much more thorough analysis.

1.4.2 Splice Site Mutations

Splice site mutations are those that either directly impact an existing splice site or create a new one elsewhere in the gene. In *DMD,* those that disrupt canonical sites can have any combination of a number of consequences in the

mature transcript: skipping of the affected exon; skipping of adjacent exons; whole intron retention; and “slippage” of the recognised splice site to a different site nearby, resulting in a partial truncation or extension of the exon (Ikezawa *et al.* 1998; Deburgrave *et al.* 2007; Zhang *et al.* 2007; Takeshima *et al.* 2010; Fletcher *et al.* 2013; Juan-Mateu *et al.* 2013). In some patients multiple transcripts can be detected that are mis-spliced to varying degrees, indicating that these mutations are “leaky” in their effect (Deburgrave *et al.* 2007; Takeshima *et al.* 2010; Juan-Mateu *et al.* 2013). It is clear that such complex splicing profiles as these could produce correspondingly unique disease phenotypes in the affected patients - phenotypes that may defy easy diagnosis and respond unpredictably to established treatments. At present, there is no detailed theory for predicting the effect of disruption to a given splice site, nor for explaining why there is so much variation in the spliceosome’s response to disruption, from one splice site to the next.

1.4.3 Pseudoexons

An unusual subset of splice site mutations, pseudoexons are segments of intronic sequence that are erroneously recognised as exons by the spliceosome, due to one or more activating mutations (see Fig. 1.4 for an example). With its enormous complement of intronic sequence, the *DMD* gene is especially prone to such mutations. To date, 45 unique *DMD* pseudoexon initiation events have been reported in the literature, across 23 *DMD* introns (Roberts *et al.* 1993; Ferlini *et al.* 1998; Ikezawa *et al.* 1998; Dwi Pramono *et al.* 2000; Suminaga *et al.* 2002; Tuffery-Giraud *et al.* 2003; Yagi *et al.* 2003; Beroud *et al.* 2004; Cagliani *et al.* 2004; Tran *et al.* 2005; Ishibashi *et al.* 2006; Deburgrave *et al.* 2007; Zhang *et al.* 2007; Gurvich *et al.* 2008; Madden *et al.* 2009; Takeshima *et al.* 2010; Khelifi *et al.* 2011; Magri *et al.* 2011; Malueka *et al.* 2012; Trabelsi *et al.* 2014; Greer *et al.* 2015). Of these 45 cases, in 17 the triggering mutation is an SNP that creates a strong acceptor or donor splice site.

While this may seem a high number, even in the *DMD* gene, pseudoexons are exceedingly rare in comparison to other splicing mutations (Aartsma-Rus *et al.*

2006). Sironi *et al.* (2004) have also pointed out that this total is far smaller than we would expect based on our current understanding of exon splicing, by which they predict approximately one pseudoexon per kilobase of intron. These authors and others (Sun and Chasin 2000) suggest that the presence of exonic silencer elements (ESEs) in the sequence surrounding the acceptor and donor splice sites are accountable for much of the discrepancy between predicted and actual pseudoexon frequency. Sun and Chasin go on to suggest that the secondary structure of the transcript, both during and after maturation, may play an additional regulatory role.

It is overwhelmingly the case that most of the mechanisms that define *DMD* splice sites are understood only in the general sense. This leaves researchers with no choice but to make, at best, educated guesses at how these mechanisms are interacting with each other at any specific site. It is therefore essential that we expand our knowledge of the spliceosome and how it interacts with the *DMD* gene transcript.

Figure 1.4: Description of a pseudoexon in intron 11 of human *DMD*. (a) RT-PCR of patient *DMD* mRNA reveals a larger isoform that includes pseudoexon 11a. (b) The position of pseudoexon 11a within the mature transcript, and the sequence of its 5' and 3' ends, are shown here. The mutation that triggered its inclusion (c.1336_1337del) is also indicated. (c) The position of pseudoexon 11a within intron 11. *Image source: Malueka* et al. *2012.*

1.5 Bioinformatics and the Dystrophinopathies

Bioinformatics software and web tools are frequently used in many fields of research into the dystrophinopathies. Utilities such as Human Splicing Finder (Desmet *et al.* 2009), Splice Aid 2 (Piva *et al.* 2012) and ESE Finder (Cartegni *et al. 2003*; Smith *et al.* 2006) permit rapid and easy calculation of sequence properties such as Shapiro-Senapathy splice site scores (Senapathy *et al.* 1990), branch points, and splicing factor binding sites. Of these three tools, ESE Finder and Human Splicing Finder offer essentially the same functionality, albeit through differently structured interfaces, whereas Splice Aid 2 is designed primarily for predicting splicing factor binding sites, and has the most comprehensive library of factor motifs.

Most often, tools such as these play a minor role in the research for which they are used, facilitating projects that are largely based within a traditional “wet” laboratory. Only a very few studies have put bioinformatics at the centre of research into *DMD* and the dystrophinopathies specifically.

Nouri *et al.* (2013) employed established techniques (MLPA and multiplex PCR) to identify ten dystrophinopathy patients with *DMD* point mutations. They then attempted to use a selection of web-based tools to classify these mutations and predict their phenotypic consequences. These tools did not allow the authors to make any definite conclusions about the mutations, however, for as they point out:

> “ …the predictions of these programs mostly do not agree with each other or with results from other analyses; therefore, the interpretation of the results should be done with caution.”

The authors also do not attempt to experimentally verify any of the predictions made with the web tools, though it is possible that they may publish such work in the future.

Legrand *et al.* (2011) had more success in their study, wherein they conducted a computational analysis of the tandem repeats that form the rod domain of dystrophin, examining the structure and surface properties of these regions.

They found that the repeats could be clustered into four groups, which together formed seven sub-domains of the rod region. The authors claimed these sub-domains aligned well with the regions of the rod domain that other researchers had predicted to bind F-actin, nNOS and lipids. Placing their research in the context of previous discoveries, they suggest that future work may constitute "docking studies" to reveal how well associated proteins bind at the sites they have predicted. The authors also point out that improving our knowledge of dystrophin sub-regions may enable a better informed approach to the development of exon-skipping strategies. Though they do not give further detail on this point, it can be inferred that understanding which sub-regions might be truncated or omitted from the protein will enable a more nuanced approach to targeting exons for AO-induced skipping.

In their 2012 report, Malueka *et al.* attempted to use the software algorithm C4.5 to generate a decision tree that would distinguish canonical *DMD* exons from pseudoexons, based on a dataset of various characteristics of the splice sites. The authors' intent was that this might illuminate defining features distinguishing one class of exons from the other. C4.5 itself has established utility, as it is cited in published research in a wide variety of fields. However, it was not an appropriate tool to use in this case, as the sample sizes (77 canonical exons and 15 pseudoexons) were too small to draw useful conclusions, and the authors do little to establish whether the analysis they performed was valid, based on the type of data. Nonetheless, the idea of using machine learning software to model exon splicing fate is an intriguing one and worthy of further investigation.

In the Malueka *et al.* (2012) paper, as in many studies that employ ready-made machine learning software or web utilities, there is a tendency to treat the tools as "black boxes" (Vihinen 2014): the data is collected and fed through, usually with default settings for most of the parameters, and the output is taken at face value without much deeper scrutiny of its validity. Nouri *et al.*'s (2013) study exemplifies another common but less severe problem with many bioinformatics studies: Although their *in vitro* work was quite rigorous, they could not derive any useful information from their *in silico* analysis because of the conflicting predictions of the various web tools they used.

The Legrand *et al.* (2011) study is certainly the most well-constructed of the three found. The aims are thoroughly contextualised and justified, and it makes testable hypotheses that will yield useful information whether they are upheld or disproven (e.g. if F-actin is found to not associate with the predicted sub-domain, this may lead to the construction of a more accurate explanatory model).

Both the positive and negative aspects of these studies indicate that there is substantial potential for further application of bioinformatical methods to understanding the *DMD* gene, its transcripts and proteins. While promising work has been done in this area, it is at present a largely unexplored field.

1.6 Conclusion

Exon-skipping antisense oligomer therapies are arguably the most promising treatment option for dystrophinopathy at this time, though there is still much work to be done towards improving their efficacy and delivery. Because these therapies are tailored, each stage of design and evaluation is underpinned by a detailed understanding of the *DMD* gene, its mutations, splicing, and the effect of AOs on the transcripts. A detailed knowledge of these aspects of *DMD* is therefore vital to the continued improvement of AO therapies and their concomitant patient outcomes as a result of therapy.

This project takes three diverse approaches to improving our understanding of *DMD* splicing and mutations:

- Chapter 3 describes the development of a PCR-based approach to finding and sequencing intronic breakpoints in DMD and BMD patients with whole exon deletions. This approach is successfully applied to seven unique *DMD* cell lines, and the natures and putative origins of their deletion junctions are discussed.
- Chapter 4 shows a simple approach to detecting rare *DMD* and *NF1* alternative splicing events in the transcript population of normal human myoblasts. The existence of these transcripts in the absence of any obviously

causative genetic variance underlines the imperfect fidelity of the spliceosome and the necessity of the nonsense mediated decay pathway.

- Chapter 5 presents a bioinformatic analysis of the relationship between splice factor motifs and successful exon-skipping antisense oligomer target sites. This analysis unexpectedly discovered evidence for a stronger link between myotonic dystrophy and *DMD* transcript mis-splicing than was previously thought to exist.

Chapter 2

General Materials and Methods

Note: A full list of all materials and their suppliers is included in Appendix E.

2.1 Standard RT-PCR Protocol

Primary PCRs were prepared in a laminar flow hood using the One-Step RT-PCR System with Platinum® Taq DNA Polymerase kit from Thermo Fisher Scientific, using 50ng of total RNA as template. They were run for cycles of 55˚C cDNA synthesis for 30 minutes, 94˚C denaturation for 2 minutes, then 35 cycles of 94˚C denaturation for 1 minute, 55˚C annealing for one minute and 68˚C extension for 2 minutes. From each primary reaction, 1uL of product was used as template for the secondary reaction that followed.

2.2 Standard Secondary PCR Protocol

Secondary PCRs were prepared in a laminar flow hood using the AmpliTaq Gold® DNA Polymerase with Buffer II and $MgCl_2$ kit from Thermo Fisher Scientific. If using primary PCR product as template, the secondary PCRs were run for cycles of 94˚C hot start for 6 minutes, then 30 cycles of 94˚C denaturation for 30 seconds, 55˚C annealing for one minute and 72˚C extension for 2 minutes. If using gDNA as template, the cycle number was increased to 35 and all other parameters were kept as described.

2.3 Cell Resurrection Protocol

2.3.1 Preparation of flasks and media

All media and buffers to be used were placed in a 37˚C water bath and allowed to warm to temperature. If the cells to be resurrected were myoblasts, the bottom of a T75 flask would be evenly coated with 3mL of matrigel and allowed to set in an incubator for 45 minutes before being cleared of the matrigel excess. Uncoated sterile T75 flasks were used for fibroblast cultures.

2.3.2 Retrieval of cells

The desired vial of cultured cells was retrieved from liquid nitrogen storage and placed in a sealable transfer vessel partially filled with wet ice. The transfer was closed and carried to the tissue culture laboratory to begin resurrection.

2.3.3 Defrosting cells

Cell vials were transferred to a 37˚C water bath for the minimum time necessary to defrost, so as to minimise the toxic effects of the cryoprotectant DMSO on the cells. Defrosted vials were moved to a laminar flow hood and their contents pipetted to 15mL Falcon tubes, each filled with 9mL of 37˚C Dulbecco's Modified Eagle Medium 5% Horse Serum, which would dilute the DMSO.

2.3.4 Centrifuging

Falcon tubes were sealed and centrifuged at 0.6rcf for 180 seconds.

2.3.5 Proliferation Media

A volume of 9mL of the appropriate proliferation media was pipetted to each prepared flask. Dulbecco's Modified Eagle Medium 15% Fetal Calf Serum (with 1x GlutMax) was used for fibroblasts, and Ham's F-10 20% Fetal Calf Serum 0.5% Chick Embryo Extract was used for myoblasts.

2.3.6 Resuspension

Excess liquid was removed from the pelleted cells and they were resuspended in 1mL of their appropriate proliferation media. The resuspension was pipetted to the prepared flask, then the flask was sealed and gently swirled to ensure an even coating of the cells on its base.

2.3.7 Incubation

Cell quality and quantity was checked by viewing the flask under a microscope. The outside of the flask was sterilised with 70% ethanol and transferred to an incubator. Newly prepared flasks were left for at least 24 hours before checking the establishment of the resurrected culture under a microscope.

2.4 Cell Culture Expansion and Maintenance Protocol

Once established, cell cultures were checked under microscope about once every two days to monitor rate of growth. Media was fully changed every three days. Cultures were propagated to double the number of equally sized flasks once they reach a confluence level of 80% or above (see section 2.5: Cell Propagation protocol).

2.5 Cell Propagation Protocol

New flasks and media were prepared as described in section 2.3.1.

2.5.1 Rinsing

For each flask to be propagated, the old media was removed and the cells were gently rinsed with 10mL of 1x PBS buffer.

2.5.2 Propagating

After removing the buffer, 3mL of 0.25% trypsin solution was added to each culture flask, and they were incubated at 37˚C. Once 4 minutes had elapsed, the flasks were inspected under a microscope to determine whether the cells had separated and were floating freely. If some remained stuck down, the flask was gently tapped on a hard surface until all had lifted off.

2.5.3 Deactivation of trypsin

The trypsin in each flask was deactivated with 7mL Dulbecco's Modified Eagle Medium 5% Horse Serum.

This cell suspension was transferred to a Falcon tube. From this point, cell propagating proceed as per cell resurrection protocol, sections 2.3.4 to 2.3.7, save that the cell resuspension was evenly divided amongst the new flasks in section 2.3.6.

2.6 Cell Transformation Protocol

2.6.1 Preparation of flasks

Flasks were coated with 3mL of Poly-D-Lysine for 1 hour at room temperature. After removing the excess Poly-D-Lysine, the flasks were then coated with 3mL of matrigel for 45 minutes at 37˚C. The excess matrigel was removed before proceeding.

2.6.2 Preparation of viral suspension

Aliquots of 10uL of MyoD adenovirus per flask were retrieved from freezer storage, defrosted, and diluted 1-in-10 in sucrose solution for a multiplicity of infection of 200.

For each fibroblast culture to be transformed, the cell-propagation protocol was followed up until the point where only the cell pellet remains in the Falcon tube. This pellet was resuspended with Dulbecco's Modified Eagle Media 5% Horse Serum, inoculated with the viral suspension, and transferred to the coated flask. The flask was then sealed and transfer to an incubator.

After 3-4 days, cells were viewed under a microscope to confirm that cell transformation had occurred and that their RNA was therefore ready to be extracted.

Chapter 3

Fractal PCR: A Strategy For Exonic Deletion Breakpoint Mapping in the DMD Gene

3.1 Introduction

As was discussed in the General Introduction, there is a diverse array of mutation types that can give rise to DMD or BMD. However, a majority of pathological lesions in the *DMD* gene are deletions (Sherrat *et al.* 1993), many of which are large enough to encompass one or more entire exons.

The clearest predictor of the phenotype a large deletion will produce is the combination of exons it encompasses. If the absence of these exons disrupts the open reading frame, the processed mRNA will either be translated into a functionless protein, or be degraded by nonsense mediated decay before translation can be completed (Hug *et al.* 2016). If, on the other hand, the deletion preserves the reading frame, then it is far more likely a functional protein of some form will be produced (Monaco *et al. 1988)*; though this cannot be guaranteed (Kesari *et al.* 2008). The *DMD* gene, and by extension the dystrophin protein, does possess a degree of redundancy, but not for all exons (van den Bergen *et al.* 2014). A truncated dystrophin protein may retain much of its structural function, but lost exons in the hinges (Nakamura *et al.* 2008), actin-binding domain (Banks *et al.* 2007) or dystroglycan-binding domain (Ishikawa-Sakurai *et al.* 2004) of the gene will necessarily impact the secondary functions of those parts of the protein, and create a phenotype distinct from that which may arise from a similarly-sized in-frame deletion elsewhere in the gene.

The effects of *DMD* whole exon deletions are not limited to the lost exon complement. Intronic sequences have important regulatory roles both pre-splicing (Li *et al.* 2012) and post-splicing (Rearick *et al.* 2011) of the mRNA, and deletions that encompass functional regions such as these are likely to have negative impacts on the phenotype - impacts that can vary greatly even among patients with identical exonic deletions (Muntoni *et al.* 2003). In some cases, the specific span of an intronic deletion may trigger the inclusion of a deleterious pseudoexon in the predominant transcript (Ferlini *et al.* 1998; Zhang *et al.* 2007; Khelifi *et al.* 2011; Greer *et al.* 2015). Given these factors, and the role that intron secondary structure plays in pre-mRNA splicing (Buratti and Baralle 2004), it is plausible that the unique intronic span of a *DMD* deletion may affect how readily the nearby exons will respond to exon-skipping AOs. For these

reasons, it is of great value to both the patient and the researcher to fully map the intronic breakpoints in *DMD* whole exon deletion cases where the genotype and phenotype show poor correlation. Not only will this knowledge empower the patient to better understand their own disease, but comparison of symptoms between patients with mapped whole exon deletions may lead to the discovery of new regions of import within the *DMD* introns, and improve insight into how and why these mutations originate.

With an average intron size in *DMD* of ~27kb and an intron size range of 107bp to over 248kb (Beroud *et al.* 2004), determination of the exonic span of a deletion in this gene usually offers only the most general clues as to intronic start and end points. At present, the options for achieving precise deletion breakpoint definition are limited. Whole Genome Sequencing (WGS) may sometimes be a viable choice at the level of an individual genome, but the cost, while lower than ever before - A$1245 as of October 2015 (National Human Genome Research Institute 2016) - is still prohibitive for some, and certainly remains so for researchers working on larger experimental populations. Gualandi *et al.* (2006) succeeded at defining breakpoints in *DMD* introns 2, 6 and 7 with the more targeted approach of long-range PCR walking, though this method has the potential to be time-intensive and reagent-consumptive: moving in steps of 2kb at a time, it could take up to a hundred sequential reactions to find a breakpoint in the largest of the *DMD* introns. Ishmukhametova *et al.* (2012) and Wiszniewska *et al.* (2014) achieved excellent results with their custom array comparative genomic hybridisation (array-CGH) chips, but at A$1000 per sample for DMD deletion/duplication analysis (Emory Genetics Laboratory 2016), the pricing for commercially available array-CGH remains prohibitive. There is a clear need for a strategy for precise breakpoint definition that is both efficient and affordable.

Presented here is a new PCR-based approach to mapping *DMD* large deletion breakpoints quickly and cost-effectively, called Fractal PCR. An analogous technique, Log-PCR, has previously been successfully demonstrated by Trimarco *et al.* (2008), though their experimental design was single-phase and focused on defining mutations only to the whole-exon level. The Fractal PCR protocol was designed with the intent of making exonic deletion analysis and

breakpoint definition possible for laboratories whose budgets may prohibit large-scale WGS. By demonstrating its efficacy on a “worst case scenario” gene such as *DMD* (i.e. extremely large intron sizes, high exon number), we show that the technique could be readily adapted to other, smaller genes with high efficiency.

3.2 Specific Materials and Methods

3.2.1 Cell Culture

Eight patient myoblast or fibroblast cell strains, with deletions of at least one whole exon in the e45-e51 region of the *DMD* gene, were selected from our cell database (see Table 3.1 for details). These strains had been cultured and expanded from patient biopsies previously, before being cryogenically stored for later use. No entirely new cell strains were initiated for the purpose of this study. We also selected a human myoblast strain not known to carry any disease-causing allele for use as a normal control. The cells were resurrected according to our standard laboratory protocol and expanded according to the cell culture expansion and maintenance protocol. Two approximately ~90% confluent T75 flasks were prepared for each cell strain: one was used for RNA extraction (using a Zymo Research Direct-zol RNA MiniPrep kit, according to kit protocol), and the other for gDNA preparation (using an Invitrogen PureLink Genomic DNA Kit, according to kit protocol).

3.2.2 Transformation of Human Fibroblasts to Myoblasts

While the preference is to use myoblasts wherever possible, since these can be differentiated into myotubes that will transcribe *DMD* and other muscle-specific genes, in some cases myoblasts were unavailable for a particular patient. In other cases, a cell strain supplied by our collaborators as “myoblasts” would prove to be predominantly fibroblasts once established in culture. In such cases, the cell strains were only usable for this study after a further step of forced myogenesis before performing the RNA extraction (see Cell Transformation Protocol, section 2.6).

3.2.3 Confirmation of mRNA Sequence

Total RNA extracted from each experimental cell strain was used as the substrate for nested RT-PCR. Existing primers were selected from the database

according to their proximity to the deleted exons in each strain. Primary and secondary PCRs were performed according to their protocols, described in General Materials and Methods. After the secondary PCR, products were fractionated on 2% agarose gels, purified using Diffinity RapidTips (according to kit instructions), and submitted to the Australian Genome Research Facility (AGRF) (http://www.agrf.org.au) for sequencing according to the AGRF protocol.

3.2.4 Primer Design

Primers were designed using NCBI Primer-Blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). The number of primer pairs to return was set at 50, to maximise the available options for primer pair selection. The melting temperature range was set to 52˚C minimum, 55˚C ideal, 58˚C maximum, (Wilton, Pers. Comm). The primer size range selected was 18nt-20nt-22nt, with a GC content range of 40%-60%, although in practice the value of the selected primers was rarely higher than 50%, due to the low GC content of most *DMD* introns. These parameters were used throughout the primer design process for this study.

The interrogated *DMD* sequence target region and predicted amplicon size were adjusted for each primer pair, according to the experimental design shown in Fig. 3.1. Amplicon size was allowed to vary up to 25bp greater or lesser than target size. For example, if the target size was 400bp, the size range would be set to 375-425bp. The region interrogated usually consisted of 500bp either side of the junction of each intron segment (eg. for an intron 80,000bp long, the first search space would be 9500-10500), though it was sometimes necessary to expand the search space if no suitable primer annealing sites could be identified on the first attempt. All other primer design parameters were left at their Primer-Blast default settings.

After each Primer-Blast search had returned primer pairs, we selected one pair from the available options according to how well that pair matched the search criteria. In many cases the sequence of at least one of the primers was adjusted slightly by adding or removing one or two bases from the ends. This was done

to ensure that both terminal nucleotides were either G or C to improve the specificity of primer annealing.

To minimise cross-amplification between primer pairs in multiplex reactions, a minimum distance of 1500bp was maintained between adjacent primer pairs/ amplicons.

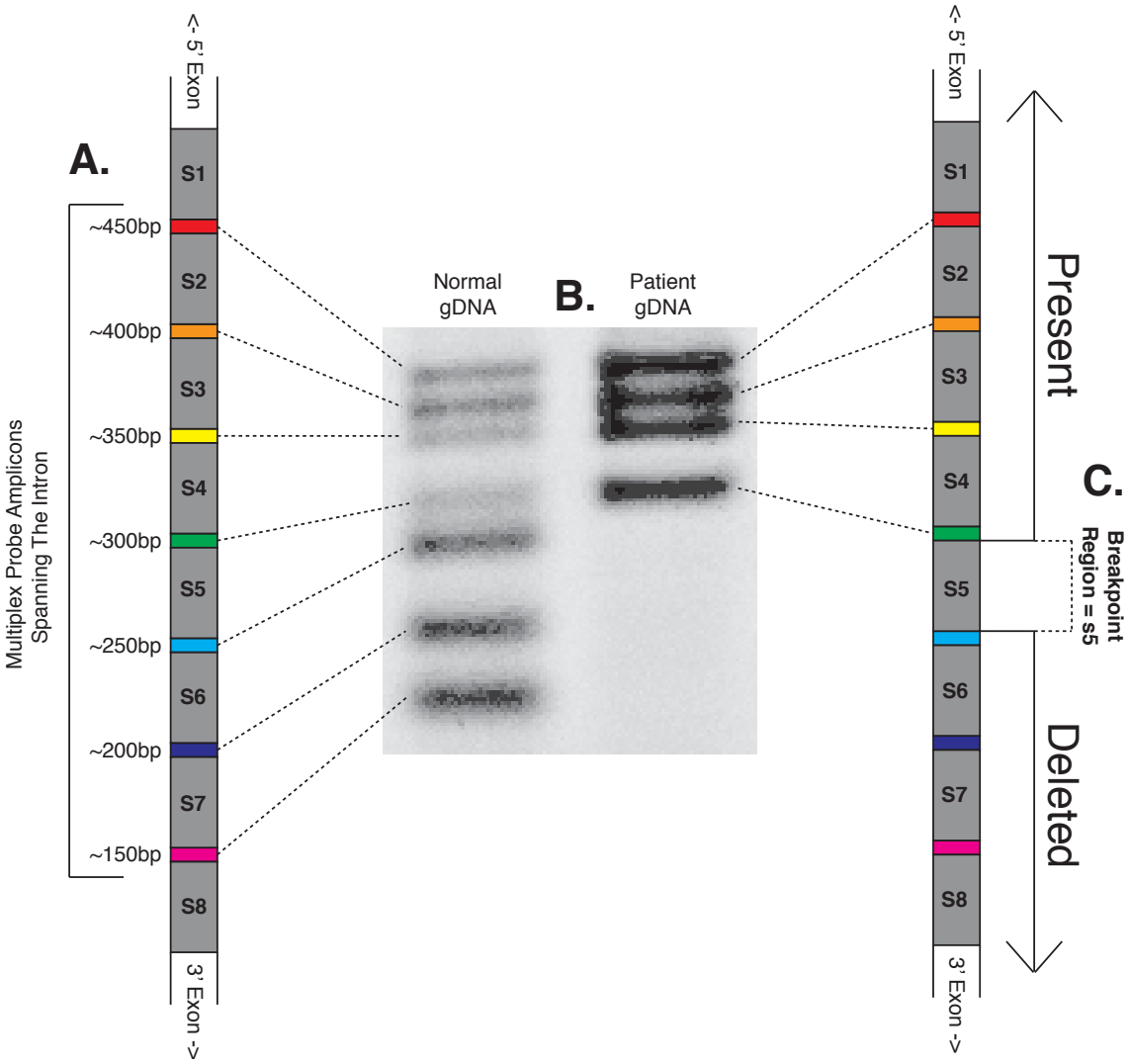


Figure 3.1: A worked example of Fractal PCR.

A. The *DMD* intron expected to carry the upstream deletion breakpoint for the patient is identified. Seven primer pairs are targeted at even intervals across this intron, dividing it up into eight sectors (S1-S8). The size of each amplicon corresponds to its position in the intron, such that a low-resolution map of the intron is created when the primers are used in a multiplex PCR.

B. Two multiplex reactions are performed using normal and patient DNA as templates. All sectors of the intron are present in the normal DNA PCR, while the patient DNA reaction shows a deletion that begins in sector 5 (S5) and extends downstream.

C. The location of the 5' breakpoint is now defined to lie within 1/8th of the entire intron (S5). A new set of primers is designed to similarly target sites within S5 and the process is repeated as necessary.

3.2.5 Primer Construction

Primers were synthesised by GeneWorks (Thebarton, Australia) at 40nM scale and “Sequencing/PCR” level purity.

In order to limit cross-annealing between primer pair products in the multiplex reaction, a minimum distance of 1.5kb was maintained between all primer pairs. This meant that many of the intron sectors being probed were only large enough to accommodate two or three primer pairs, rather than the default full set of seven.

3.2.6 Primer Verification

Upon receipt, primers were resuspended in ddH_2O, and diluted aliquots were made with a target concentration of 100ng/uL.

Primer pairs were tested individually on normal human genomic DNA (gDNA), using the standard secondary PCR protocol (section 2.2 - see Appendix D for primer sequences). Primers were dosed at 140ng per reaction to ensure a visible product. PCR products were separated on 2% agarose-TAE gels at ~98V for approximately 80 minutes and stained using Biotium GelRed™.

Any primer pairs that failed to yield a single, defined product of the expected size were retested separately. If the PCR failed a second time, this primer pair was excluded from further use in this study.

Suitable primer pairs were evaluated in a second round of verification. Using their previously validated concentrations as a starting point, each set of seven or fewer primer pairs targeting the *DMD* introns under investigation were run in a single multiplex reaction, using normal human gDNA as the template. Reaction products were separated on a 2% agarose-TAE gel and assessed for absence of cross-annealing between primer pairs (“fidelity”) and comparable product intensity within each set.

In cases where an off-target product was observed, this multiplex set underwent an elimination test, wherein the multiplex reaction was repeated multiple times with a different single primer pair removed each time. This identified which of

the primer pairs had cross-reacted, and one of the involved pairs would be excluded from the multiplex henceforth.

Once the fidelity and consistency of the multiplex sets had been established, they passed to the concentration-balancing stage. During this stage, the proportions of the primers were adjusted such that all products were of similar intensity when visualised on a gel. Upon completion of this stage, the multiplex primer sets were considered to be fully verified and ready for experimental use. For ease of pipetting and future use, the primers were combined into bulk quantities of “primer multimixes” in the determined proportions.

3.2.7 Exonic Deletion Breakpoint Mapping - Phase I

For the first phase of exonic deletion breakpoint mapping, the extracted gDNAs from each *DMD* exonic deletion patient were used as template for two reactions, one containing each multiplex set targeting the introns flanking the known exonic deletion, i.e. the region where an exonic deletion breakpoint was expected to be found. For each multiplex set, a reaction using normal human gDNA was run as a positive control. The results of this phase determined which sector of each intron spanned that patient’s exonic deletion breakpoint. This information was used to proceed to Phase II.

3.2.8 Exonic Deletion Breakpoint Mapping - Phase II

The procedures for phase II were much the same as for phase I, except that the primers were designed to target breakpoint-containing segments of an intron, rather than the entire intron (see Fig 3.1, part C). The results from this phase further narrowed the investigation region and were interpreted to inform the construction of primers for phase III.

3.2.9 Final delimitation and sequencing

Following Phase II, each breakpoint region had been narrowed to a ~2-4kb zone. Individual primer pairs were targeted across these regions to further delimit them to ~500bp, and the closest primers either side of the junction were used to amplify across the breakpoints.

Breakpoint amplicons were separated on 2% agarose gels and purified using Diffinity RapidTips, according to the manufacturer's instructions and submitted to the AGRF (http://www.agrf.org.au) for sequencing. Comparison of the sequencing data to the reference *DMD* sequence (RefSeqID: NC_000023.11) then permitted precise identification of the exonic deletion breakpoint junction site. Sequence images were created by visualising sequence data using the software program "4peaks" (Nucleobytes) and taking screen captures.

3.3 Results

Cell Strain ID	Assigned Exonic Deletion	Assigned phenotype	Cell type	Origin
1148	47-51	Duchenne Muscular Dystrophy	Myoblast	Hammersmith Hospital, London, United Kingdom
1800	45-50	Duchenne Muscular Dystrophy	Fibroblast	The Children's Hospital at Westmead, Sydney, Australia
2033	48-50	Duchenne Muscular Dystrophy	Myoblast	Dubowitz Neuromuscular Centre, London, United Kingdom
2037	48-49	Becker Muscular Dystrophy	Myoblast	Dubowitz Neuromuscular Centre, London, United Kingdom
2054	45-47	Becker Muscular Dystrophy	Myoblast	Dubowitz Neuromuscular Centre, London, United Kingdom
2055	45-49	Becker Muscular Dystrophy	Myoblast	Dubowitz Neuromuscular Centre, London, United Kingdom
2057	45-47	Becker Muscular Dystrophy	Myoblast	Dubowitz Neuromuscular Centre, London, United Kingdom
2942	51	Duchenne Muscular Dystrophy	Fibroblast	Dystrophy Annihilation Research Trust Centre, Bengaluru, India

Table 3.1: Details of the eight *DMD* exonic deletion cell strains used in this study.

3.3.1 Confirmation of DMD *mRNA sequences*

3.3.1.1 DMD mRNA sequencing for cell strain 1148 (reported del. e47-51)

Cell strain 1148 was reported as having a deletion of *DMD* exons 47-51, however, sequencing of the mRNA (Fig 3.2) revealed that the full deletion span at the mRNA level was exons 46-51.

3.3.1.2 DMD mRNA sequencing for cell strains 1800 (reported del. e45-50) and 2033 (reported del. e48-50)

Sequencing for the *DMD* mRNAs of cell strains 1800 (Fig. 3.3) and 2033 (Fig. 3.4) confirmed their reported deletions of exons 45-50 and exons 47-51 respectively.

3.3.1.3 DMD mRNA sequencing for cell strain 2037 (reported del. e48-49)

While the 2037 cell strain was provided with a genomic diagnosis of an exon 48-49 deletion, the *DMD* mRNA sequencing (Fig. 3.5) shows unequivocally that both exons are present, as are the adjacent exons. The only deviation from the reference sequence detected was a C>A SNP at the third-to-last base of exon 48, which would cause a glutamine-to-lysine substitution in the protein and have a slightly negative effect on the donor splice score (89.73 > 88.77 Shapiro-Senapathy score).

3.3.1.4 DMD mRNA sequencing for cell strains 2054 (reported del. e45-47), 2055 (reported del. e45-49) and 2057 (reported del. e45-47)

Sequencing for the *DMD* mRNAs of cell strains 2054 (Fig. 3.6), 2055 (Fig. 3.7) and 2057 (Fig. 3.8) confirmed their reported deletions of exons 45-47, exons 45-49 and exons 45-47 respectively.

3.3.1.5 DMD mRNA sequencing for cell strain 2942 (reported del. e51)

The RNA from patient cell strain 2942 (reported del. e51) did not yield any detectable *DMD* transcripts, and therefore no cDNA sequence, although the

RNA extraction did yield RNA that was of acceptable quality according to the 260/280 absorption ratio of ~2.0. These cells were cultured as fibroblasts before being transformed into myoblasts by ADV MyoD, and it may be that the forced myogenesis was inefficient, despite the fact that an examination of transfected cell morphology at time of harvest indicated they were transformed. However, the genomic deletion for this cell strain was later confirmed via genomic DNA sequencing (Fig. 3.19).

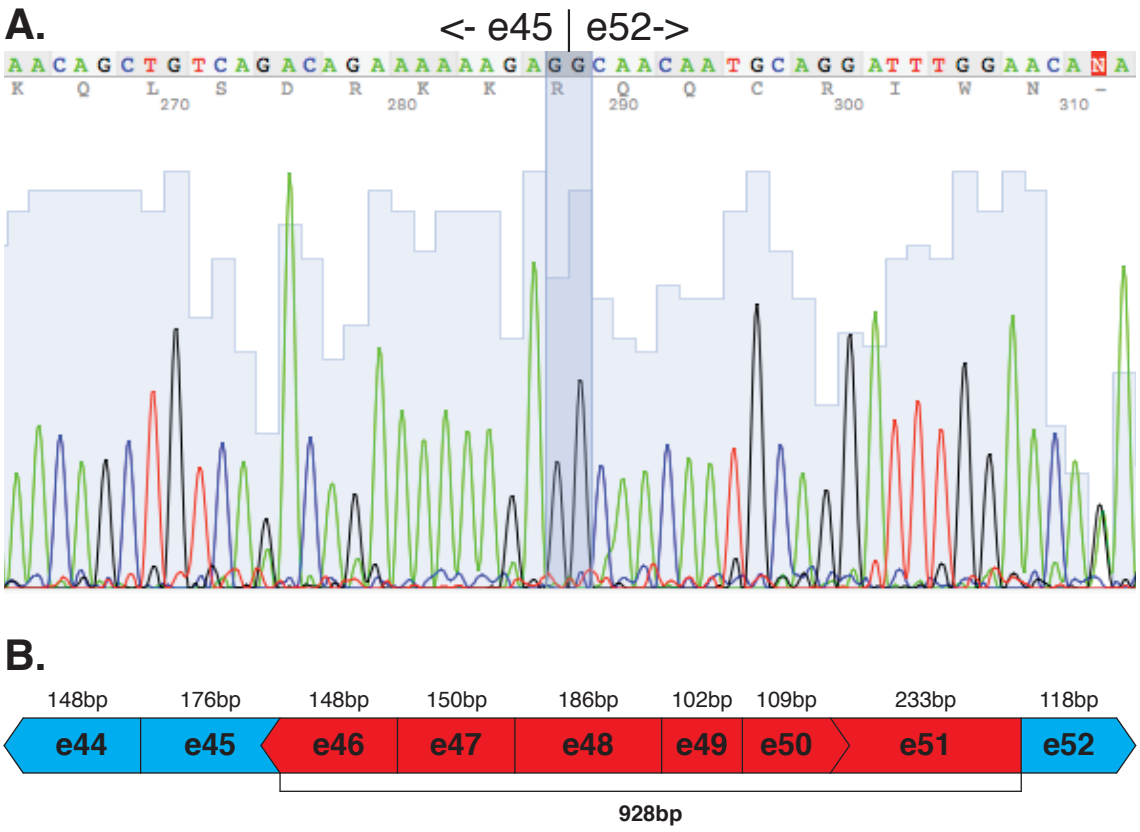


Figure 3.2: Sequencing of the RT-PCR product for DMD patient 1148 RNA (reported del. e47-51). (A) This sequencing demonstrated the additional absence of exon 46 at the mRNA level, making the total deletion span e46-51. The exon 45:52 junction is indicated. (B) The exonic span of the deletion in the mature transcript.

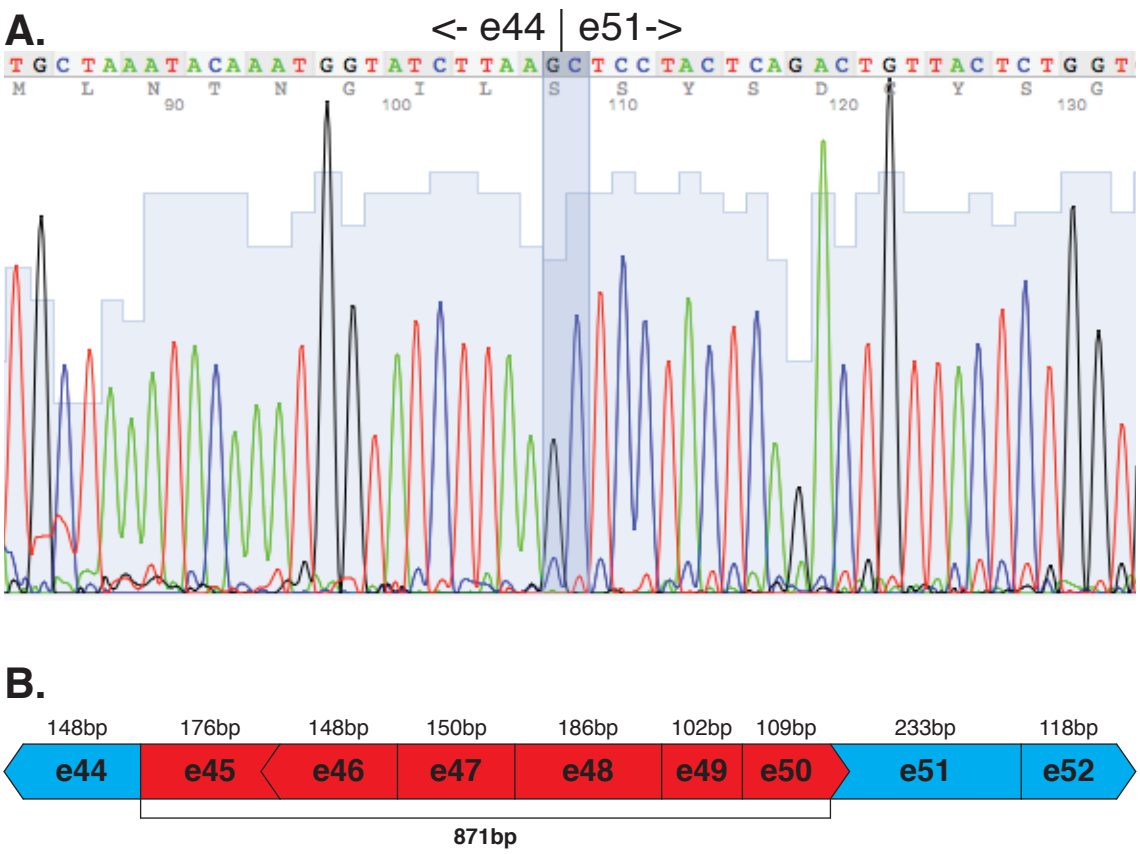


Figure 3.3: Sequencing of the RT-PCR product for DMD patient 1800 RNA (reported del. e45-50). (A) This sequencing supports genomic deletion of exons 45-50 in the mRNA. The exon 44:51 junction is indicated. (B) Exonic span of the deletion in the mature transcript.

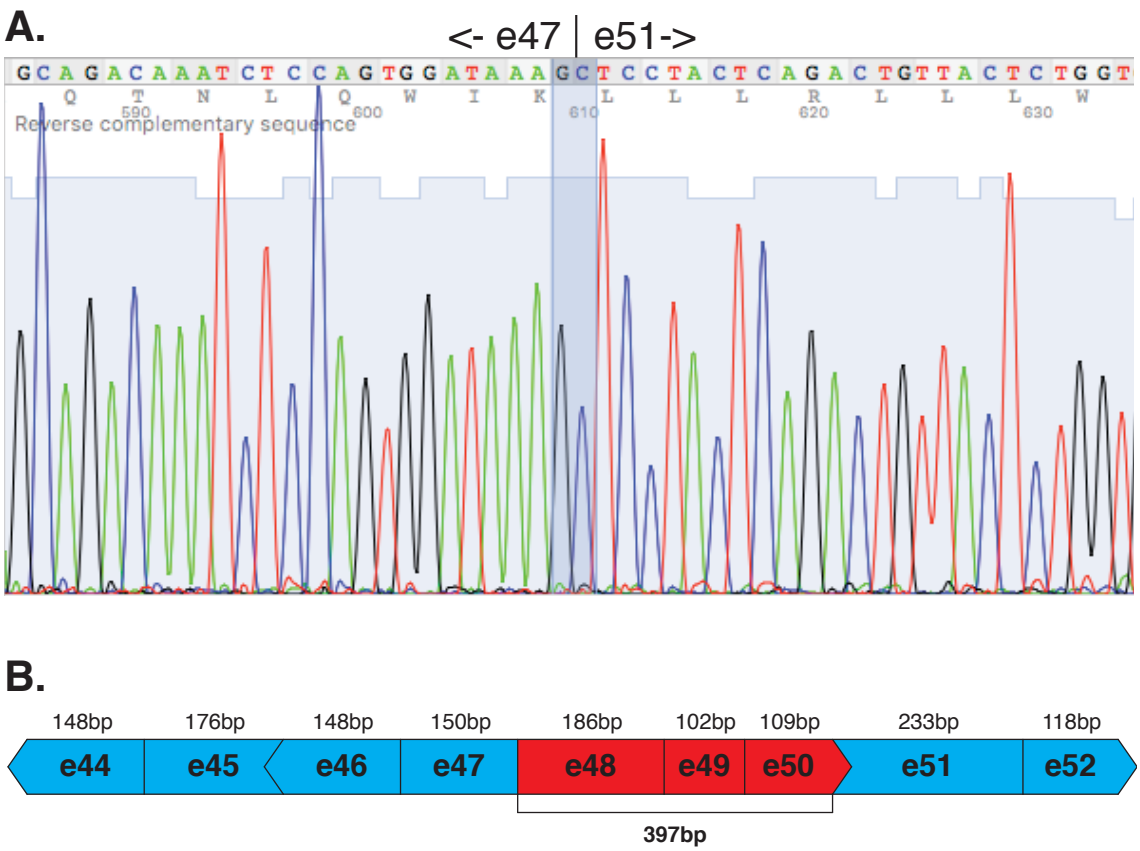


Figure 3.4: Sequencing of the RT-PCR product for DMD patient 2033 RNA (reported del. e48-50). This sequencing supports genomic deletion of exons 48-50 in the mRNA. The exon 47:51 junction is indicated. (B) The exonic span of the deletion in the mature transcript.

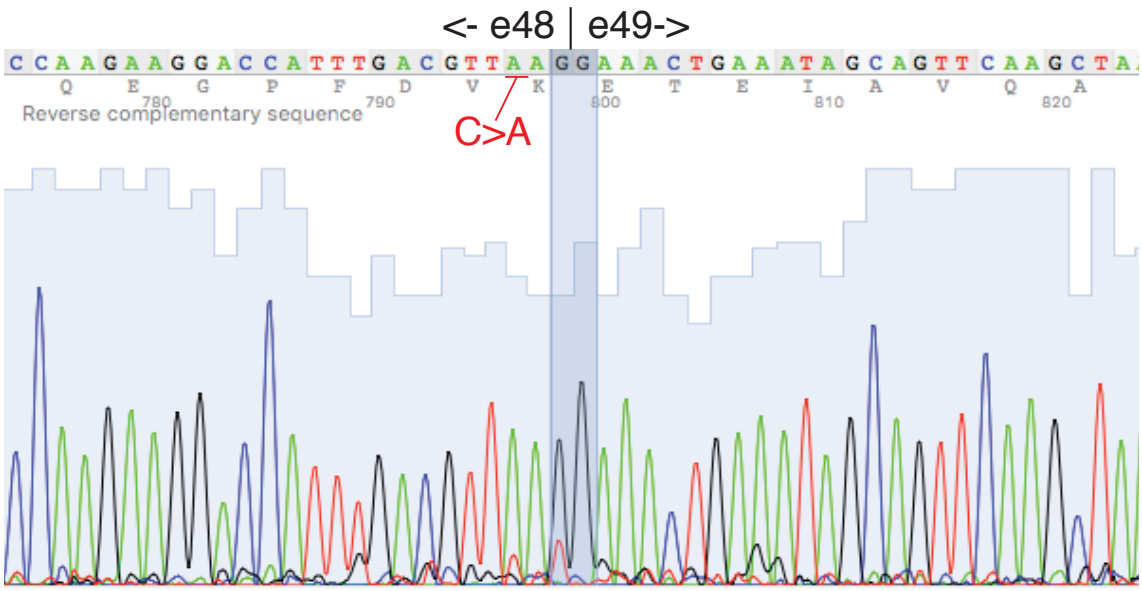


Figure 3.5: Sequencing of the RT-PCR product junction for BMD patient 2037 RNA (reported del. e48-49). All exons were present in the sequenced region. The exon 48:49 junction is indicated. One C>A SNP was detected at the third-last base of exon 48.

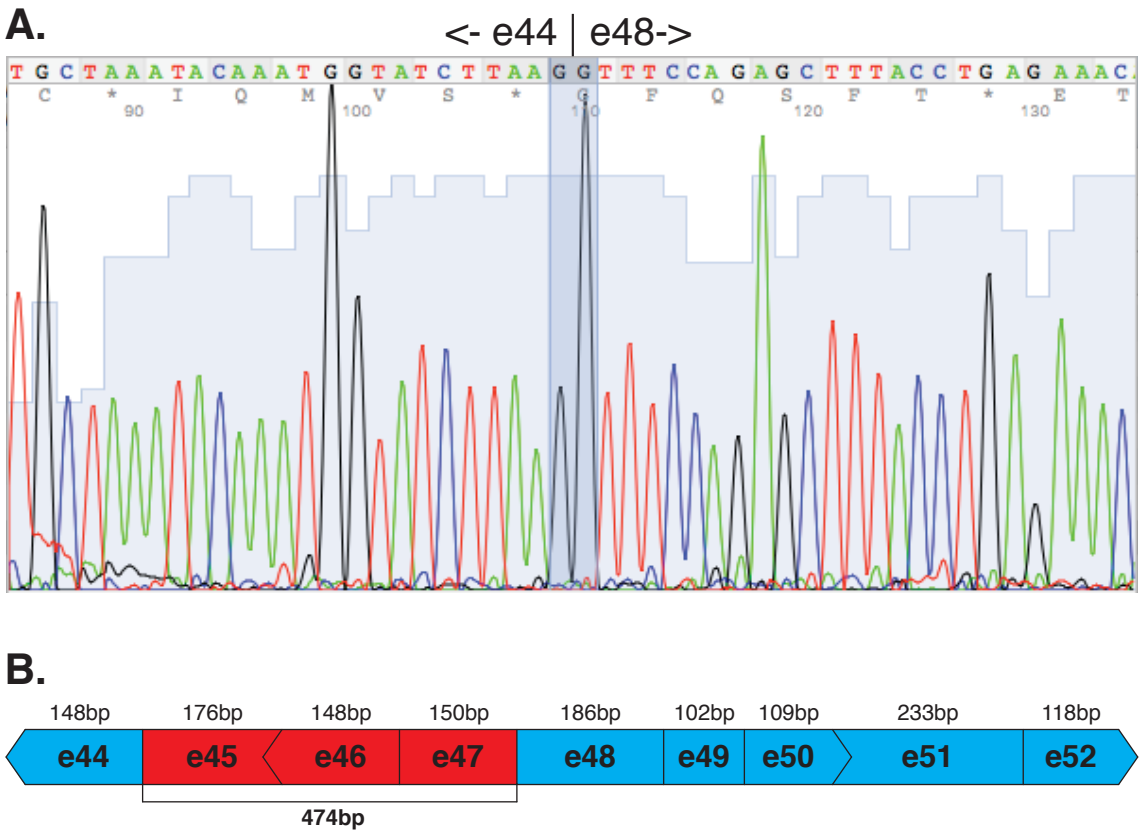


Figure 3.6: Sequencing of the RT-PCR product for DMD patient 2054 RNA (reported del. e45-47). (A) This sequencing supports the genomic deletion of exons 45-47 at the RNA level. The exon 44:48 junction is indicated. (B) The exonic span of the deletion in the mature transcript.

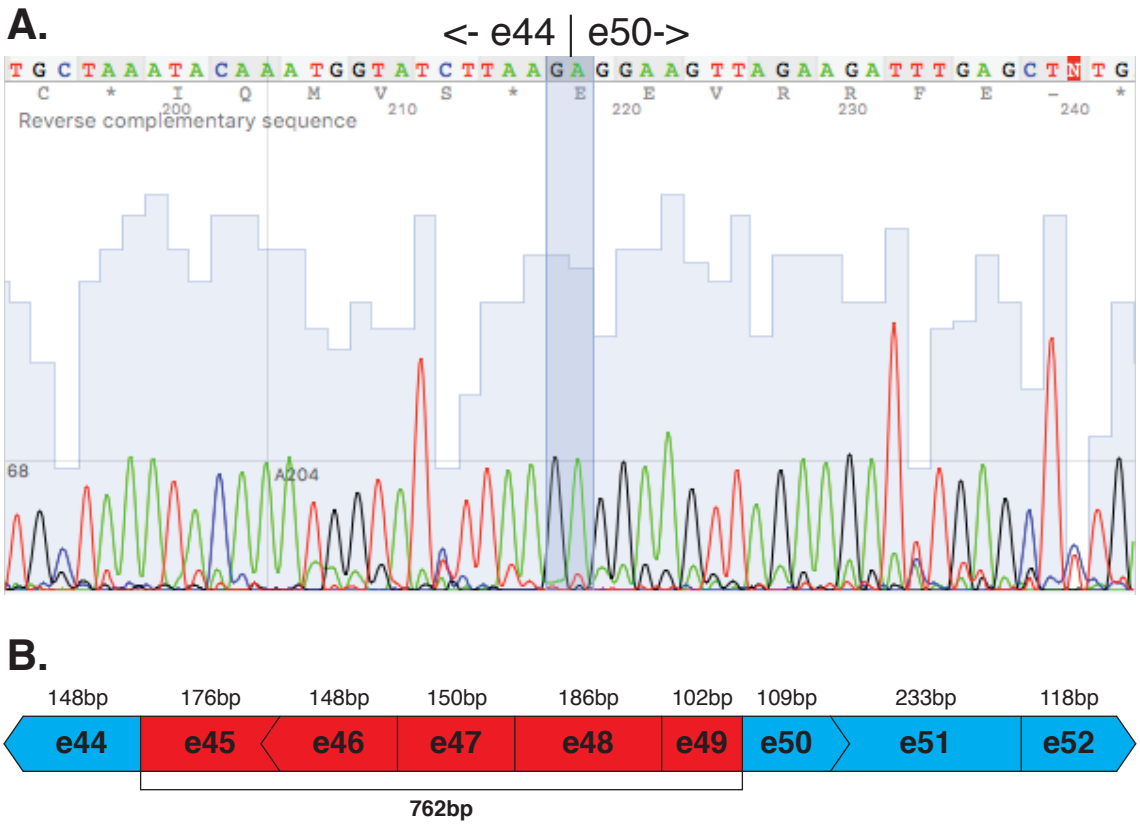


Figure 3.7: Sequencing of the RT-PCR product junction for DMD patient 2055 RNA (reported del. exons 45-49). (A) This sequencing supports the genomic deletion of exons 45-49 at the RNA level. The exon 44:50 junction is shown. (B) The exonic span of the deletion in the mature transcript.

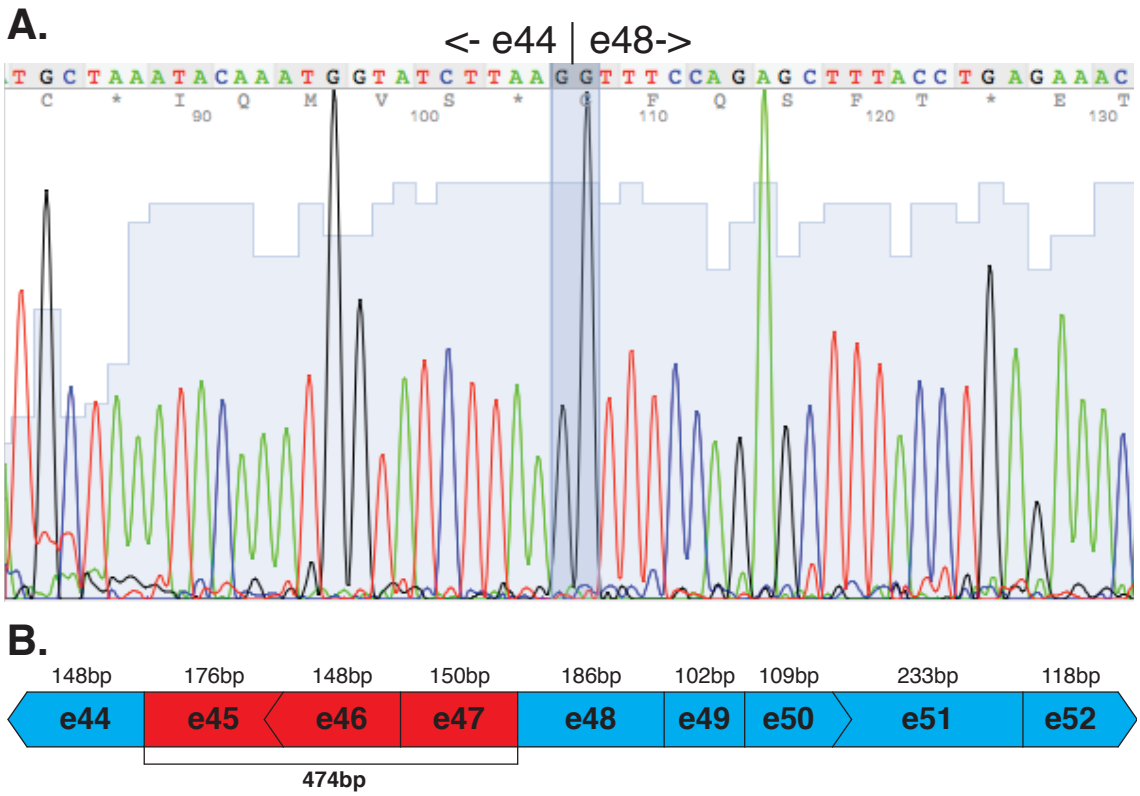


Figure 3.8: Sequencing of the RT-PCR product for DMD patient 2057 RNA, (reported del. e45-47). (A) This sequencing supports the genomic deletion of exons 45-47 at the RNA level. The exon 44:48 junction is shown. (B) The exonic span of the deletion in the mature transcript.

3.3.2 *Evaluations of Phase I exonic deletion breakpoint mapping primers*

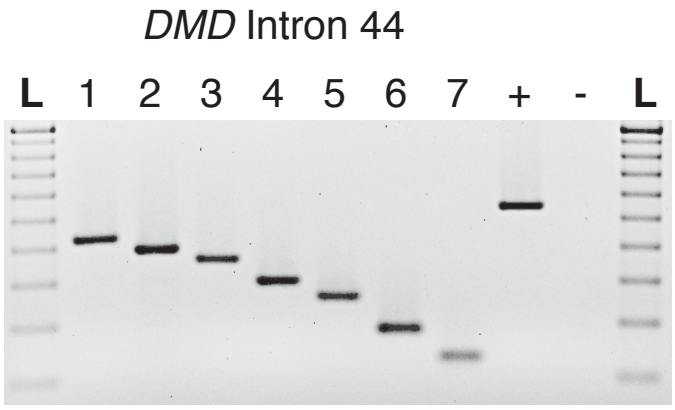
A representative example of individual primer pair evaluations is given in Fig. 3.9. All other individual primer pair evaluations are shown in Appendix A, Figs. A.3.1 to A.3.8.

3.3.3 *Phase I exonic deletion breakpoint mapping*

Final, balanced versions of all Phase I multiplex primer sets are demonstrated in Fig. 3.10 and are used for *DMD* breakpoint delimitation in Figs. 3.11a and 3.11b.

3.3.3.1 *Cell strain 2037 DNA*

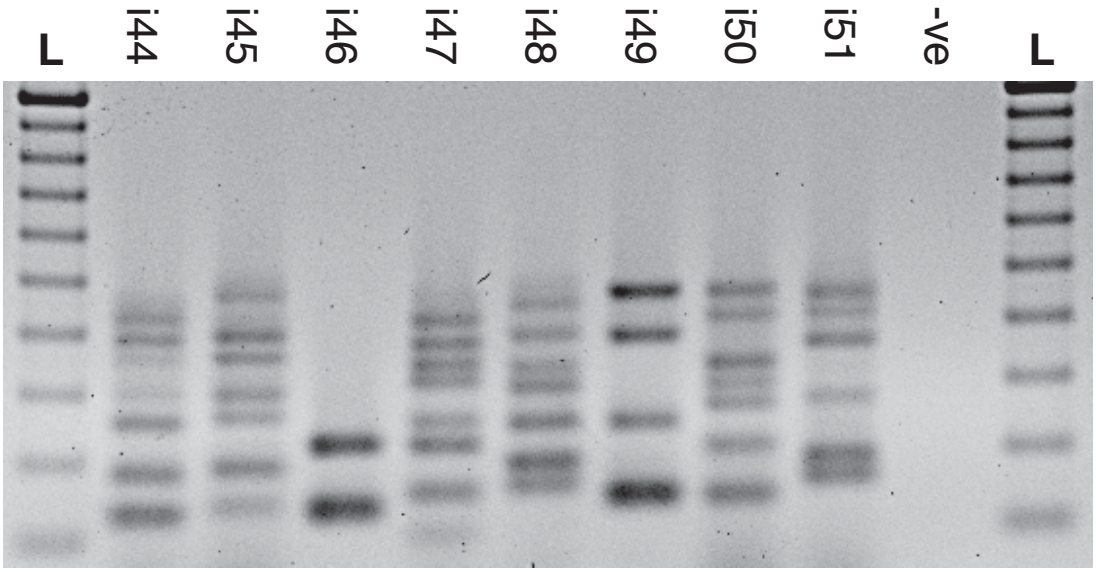
Cell strain 2037 (Fig 3.11b, grey text) did not appear to carry the reported exonic deletion of exons 48-49, or indeed any large deletion within this region (see Fig. 3.5) and no further work was done with this DNA.



Lane	F Primer	R Primer	Size (bp)
1	5867	5868	**423**
2	5869	5870	**399**
3	5871	5872	**365**
4	5873	5874	**303**
5	5875	5876	**261**
6	5877	5878	**192**
7	5879	5880	**140**
+	5742	5743	**557**
-	None	None	**0**

Figure 3.9: Evaluation of intronic primer pairs targeting *DMD* intron 44, using normal human gDNA as template. Amplicons are ordered according to the position of the intronic target site, 5' to 3' left to right. Positive control primers amplify from an easily amplified site in intron 47. "L" indicates 100bp ladder size standard.

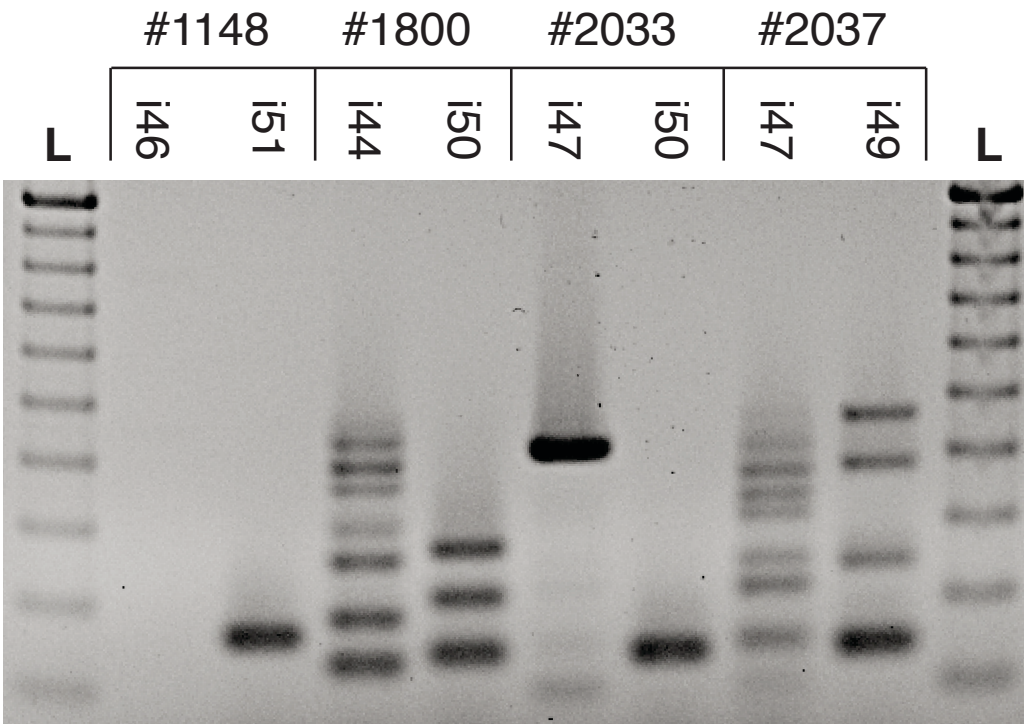
Note: See Appendix A for a full record of intronic primer pair evaluations.



Phase I Multiplex PCRs - Multiplex PCR size guide

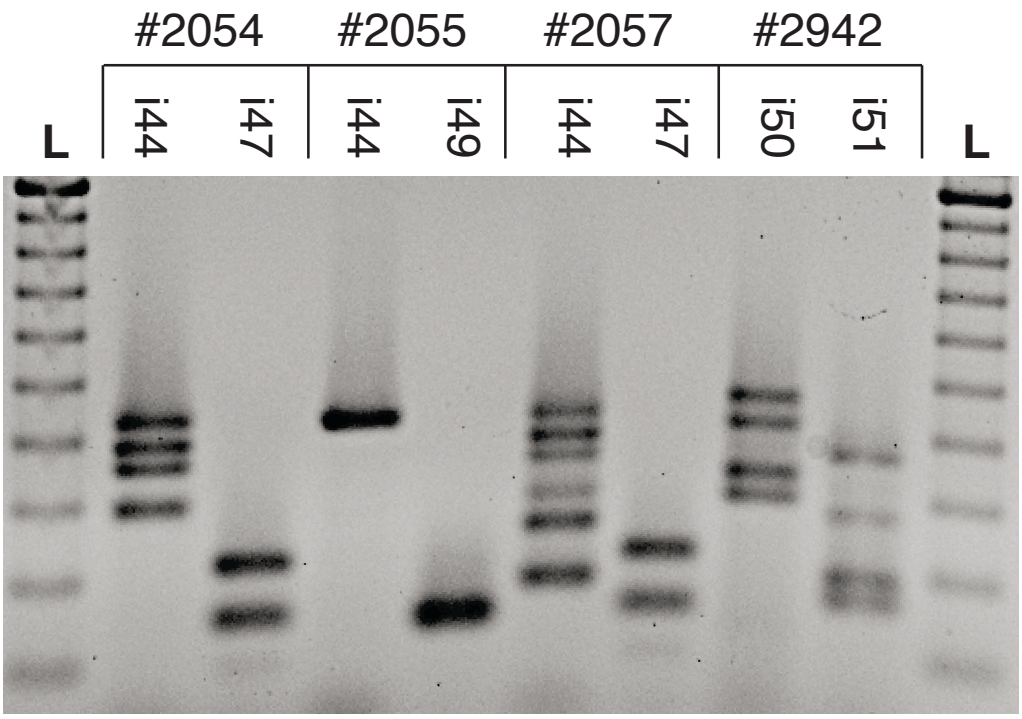
i44	i45	i46	i47	i48	i49	i50	i51	-ve
423	474	-	421	453	470	474	453	-
399	396	-	379	389	389	416	423	-
365	362	-	342	336	329	336	369	-
303	300	-	315	306	-	296	282	-
261	271	226	254	253	246	274	248	-
192	194	176	224	197	174	212	199	-
140	145	141	157	167	149	150	172	-

Figure 3.10: Final verification of *DMD* intronic multiplex primer sets showing product size guide. "L" indicates 100bp ladder size standard. During successive rounds of balancing of the multiplex primer sets, it was necessary to remove four primer pairs from the eight cocktails. The expected amplicon sizes for the excluded primer pairs are shown in grey text. The second primer pair from the i46 set (176bp) was removed as it was found to interfere with the adjacent first and third pairs. The third primer pair from the i49 set (329bp) was removed to eliminate unwanted cross-amplification with the fourth pair (246bp). The fifth pair of the i49 set and the fifth pair of the i51 set were removed because they yielded interfering off-target products with other primers in their respective multiplex reactions.



Cell Line	5' region	5' break	PhII Target	3' region	3' break	PhII Target
1148	i46	s1 of 4	i46s1	i51	s6 of 7	i51s6
1800	i44	s8 of 8	i44s8	i50	s5 of 8	i50s5
2033	i47	s2 of 8	i47s2	i50	s7 of 8	i50s8
2037	i47	s8 of 8	i47s8	i49	s1 of 8	i49s1

Figure 3.11a: Phase I intronic multiplex PCRs used to identify the *DMD* exonic deletion breakpoint regions for four cell strains, showing regions to be targeted in Phase II. "L" indicates 100bp ladder size standard. Cell strain 2037 (grey text) was identified as not bearing any deletion at its designated breakpoint (i.e. exons 48-49), via sequencing of this region of the *DMD* mRNA (see Fig. 3.5).



Cell Line	5' region	5' break	PhII Target	3' region	3' break	PhII Target
2054	i44	s5 of 8	i44s5	i47	s6 of 8	i47s6
2055	i44	s2 of 8	i44s2	i49	s4 of 5	i49s4
2057	i44	s7 of 8	i44s7	i47	s6 of 8	i47s6
2942	i50	s5 of 8	i50s5	i51	s2 of 7	i51s2

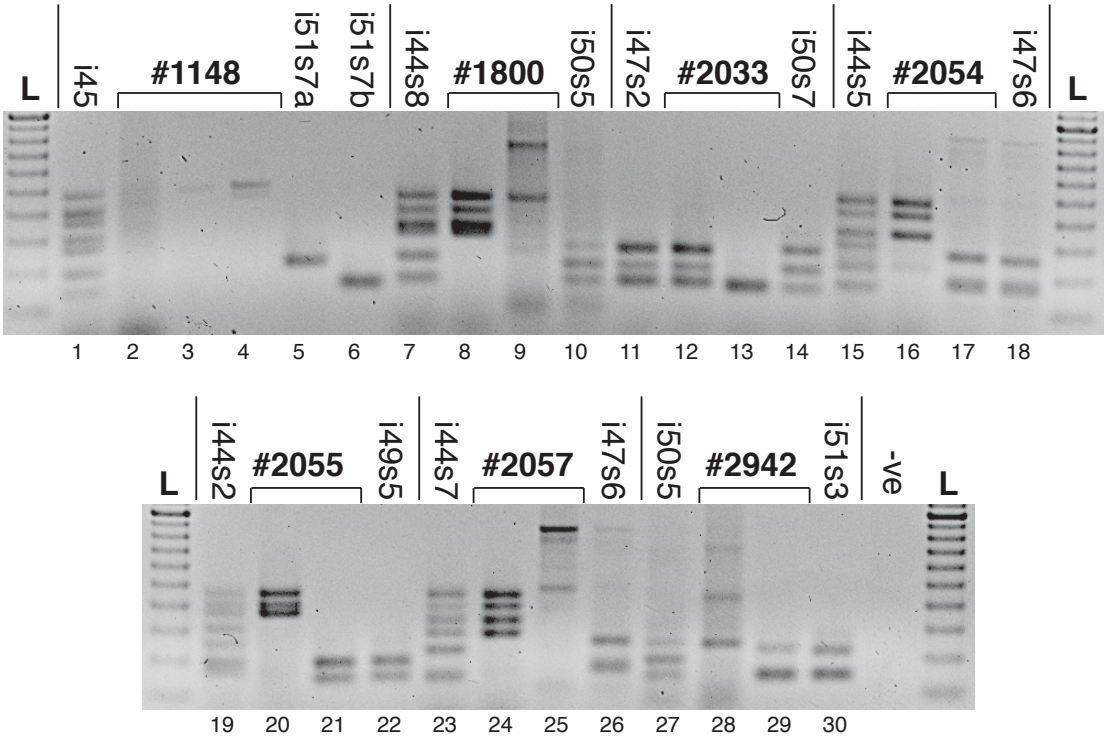
Figure 3.11b: Phase I intronic multiplex PCRs used to identify the *DMD* exonic deletion breakpoint regions for four cell strains, showing regions to be targeted in Phase II. "L" indicates 100bp ladder size standard. The products generated by each multiplex set determine which intron sector bears the breakpoint, and therefore which sector will serve as the target for the Phase II primer sets.

3.3.4 Evaluations of Phase II exonic deletion breakpoint mapping primers

Individual primer pair evaluations are shown in Appendix A, Figs. A.3.9 to A. 3.21. Phase II primer evaluations were run as a single reaction set with a single positive and negative control. These are shown in the last gel image of the series (Fig. A.3.21).

3.3.5 Phase II exonic deletion breakpoint mapping

The fifth product of the i44s5 positive control (Fig 3.12) was very faint, though this was not diagnostically important in this case. Lanes 20 and 24 illustrate that the deletion of target sites for some of the primer pairs in the multiplex often produces much more vivid products for the primer pairs whose target sites remain, presumably because these pairs have less competition for the resources of the reaction (dNTPs, enzyme etc.). In some cases (lane 9, lane 25) the complete absence of target sites leads to the production of off-target products. As shown in the size guide, three primer pairs were removed from the final versions of the cocktails (highlighted in grey). One of these pairs (i44s1, 442bp) was excluded due to production of an off-target product during the first verification trial (see Appendix A, Fig. A.3.9). The fifth pair for i44s8 (251bp) was also excluded because it created interfering off-target products with other primers in its multiplex reaction. With two intronic breakpoints per cell strain (5' and 3'), there were 13 *DMD* breakpoint zones (subsections of the sectors from phase I) identified from these results. Based on the mRNA sequencing results for cell strain 1148 (reported del. e47-51) (see Fig. 3.2), it was hypothesised that the upstream breakpoint of this cell strain may have been within intron 45. Hence, the i45 set was used for this DNA (Fig. 3.12, lanes 1 and 2) - and as can be seen in lane 2, nearly the entirety of this intron was deleted also. The two pairs of primers in the i51s7 set were created off-target amplicons during their testing phase and needed to be run in separate reactions (Fig. 3.12, lanes 5 and 6).



Size Guide

i44s1	i44s2	i44s5	i44s7	i44s8	i47s2	i47s6	i49s5	i50s5	i50s7	i51s3	i51s7a	i51s7b
442	462	459	450	441	-	-	-	-	-	-	-	-
388	409	404	398	376	-	-	-	-	-	-	-	-
324	374	331	335	325	-	-	-	-	-	-	-	-
307	314	286	288	298	-	-	-	-	-	-	-	-
276	255	245	234	251	249	265	-	252	255	-	-	-
212	201	213	222	218	199	192	198	298	195	225	201	-
158	175	164	150	150	158	174	155	158	151	156	-	143

Cell Line	5' region	5' break	PhIII Target	3' region	3' break	PhIII Target
1148	i45	s1 of 8	i45s1	i51s7	z1 of 3	i51s7z1
1800	i44s8	z5 of 7	i44s8z5	i50s5	z1 of 4	i50s5z1
2033	i47s2	z4 of 4	i47s2z4	i50s7	z3 of 4	i50s7z3
2054	i44s5	z4 of 8	i44s5z4	i47s6	z1 of 3	i47s6z1
2055	i44s2	z4 of 8	i44s2z4	i49s5	z1 of 3	i49s5z1
2057	i44s7	z5 of 7	i44s7z5	i47s6	z1 of 3	i47s6z1
2942	i50s5	z2 of 4	i50s5z2	i51s3	z1 of 3	i51s3z1

Figure 3.12: Phase II mapping of intronic breakpoints for seven *DMD* exonic deletion cell strains, showing regions to be targeted in next delimitation. "L" indicates 100bp ladder size standard. Off-target products were observed in lanes 9 and 25. Three primer pairs were removed from the final versions of the cocktails (grey text in size guide). One of these pairs (i44s1, 442bp) was excluded due to production of an off-target product during the first verification trial (see Fig. 3.14). The sixth pair for i44s7 (222bp) was removed due to visual interference with the fifth pair (234bp). The fifth pair for i44s8 (251bp) was removed for creating interfering products with other primers in the multiplex reaction. The two pairs of primers in the i51s7 set created off-target amplicons during the testing phase and needed to be run in separate reactions (lanes 5 and 6).

3.3.6 Exonic deletion breakpoint sequencing

Sequencing of the deletion breakpoint regions are shown for seven *DMD* cell lines are shown in Figs. 3.13 to 3.19. Microhomologies (short consensus sequences at at the junctions of the 5' and 3' intron fragments) are indicated where they occur, as are other sequence anomalies.

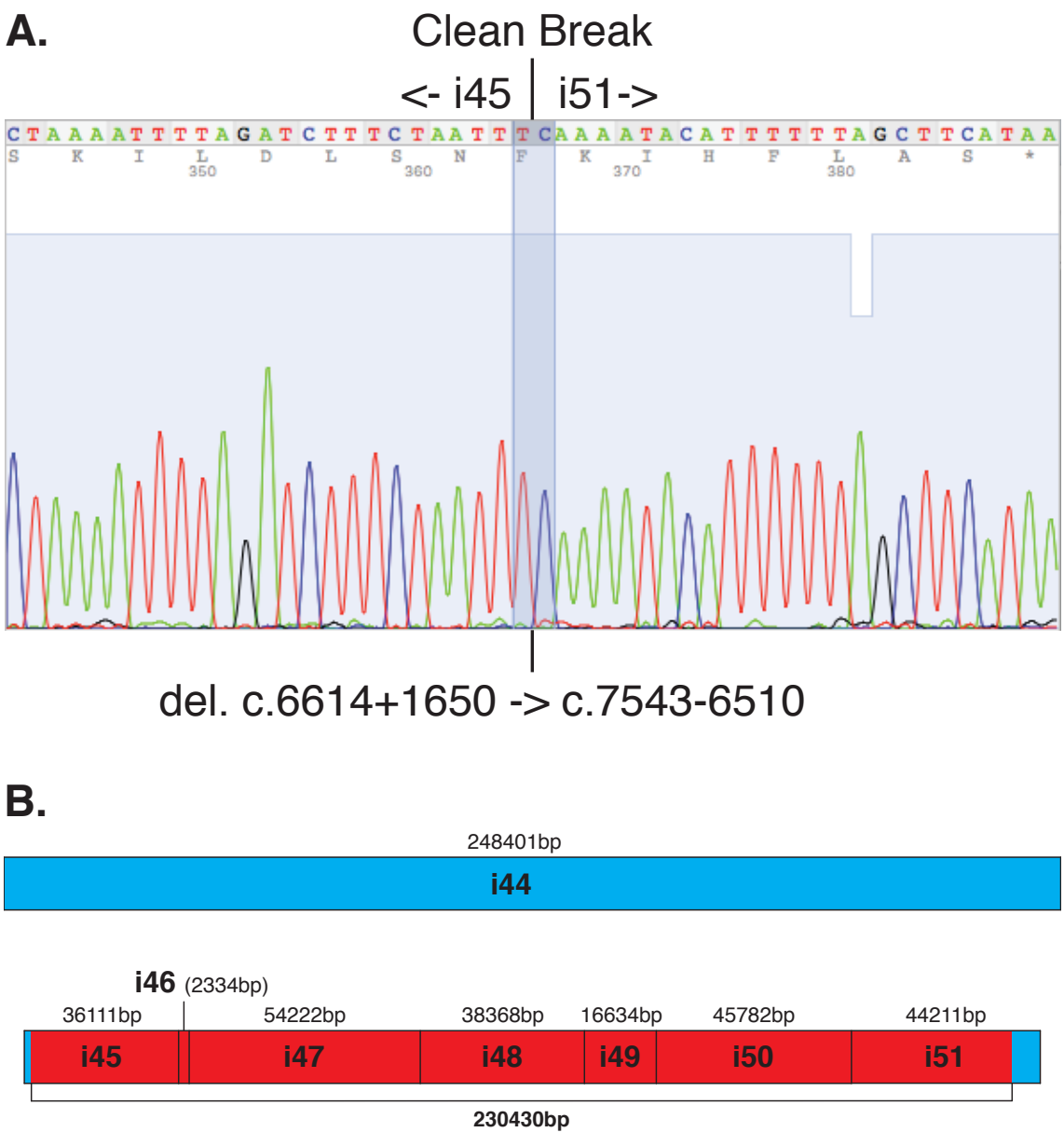


Figure 3.13: *DMD* gDNA exonic deletion breakpoint sequencing for cell strain 1148 (reported del. e47-51). (A) Though this patient was diagnosed as having an e47-51 deletion, this sequencing shows that the true span of the deletion is e46-51. (B) The intronic span of the deletion is highlighted in red. Intron and deletion sizes are also indicated.

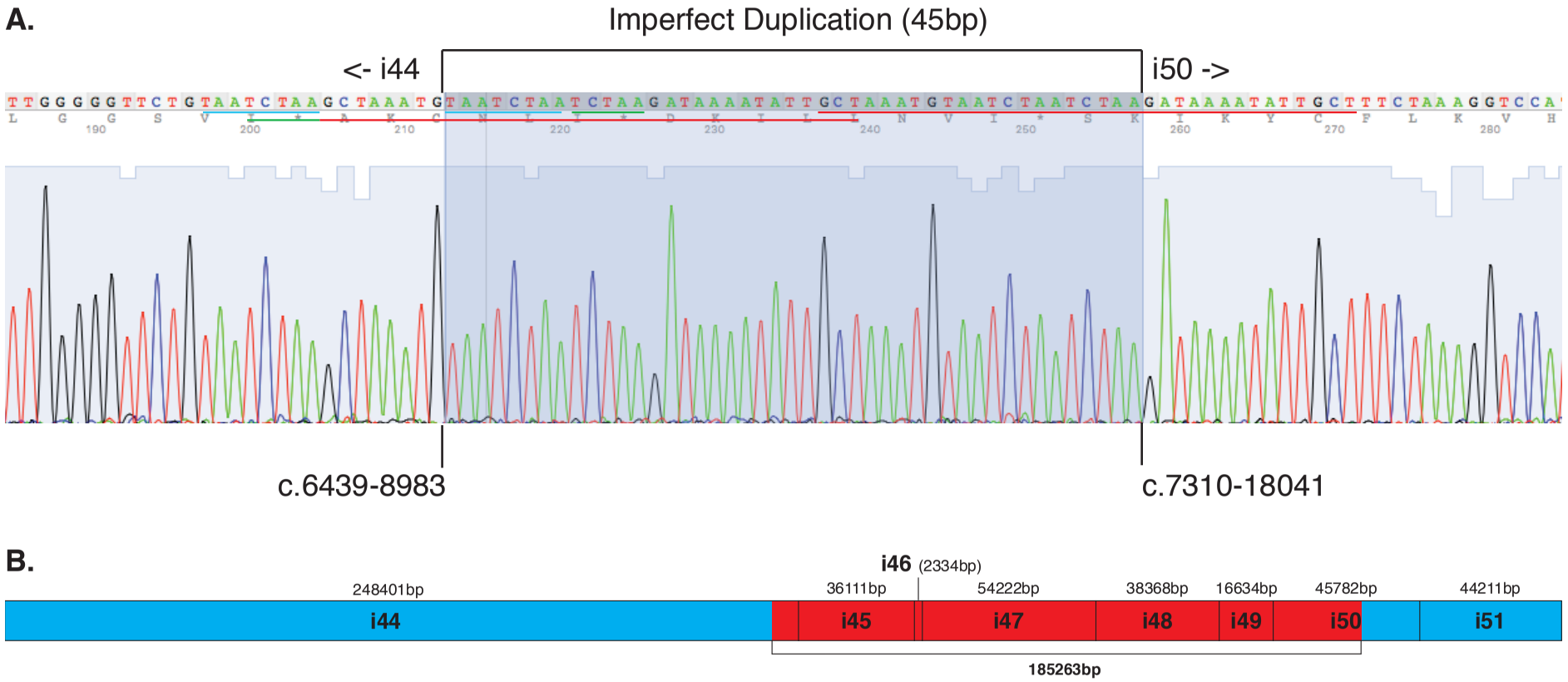


Figure 3.14: *DMD* gDNA exonic deletion breakpoint sequencing for cell strain 1800 (reported del. e45-50). (A) This DNA showed an unusual complex rearrangement (homologous sequences underlined in red, blue and green) creating 45bp of new sequence at the junction point. (B) The intronic span of the deletion is highlighted in red. Intron and deletion sizes are also indicated.

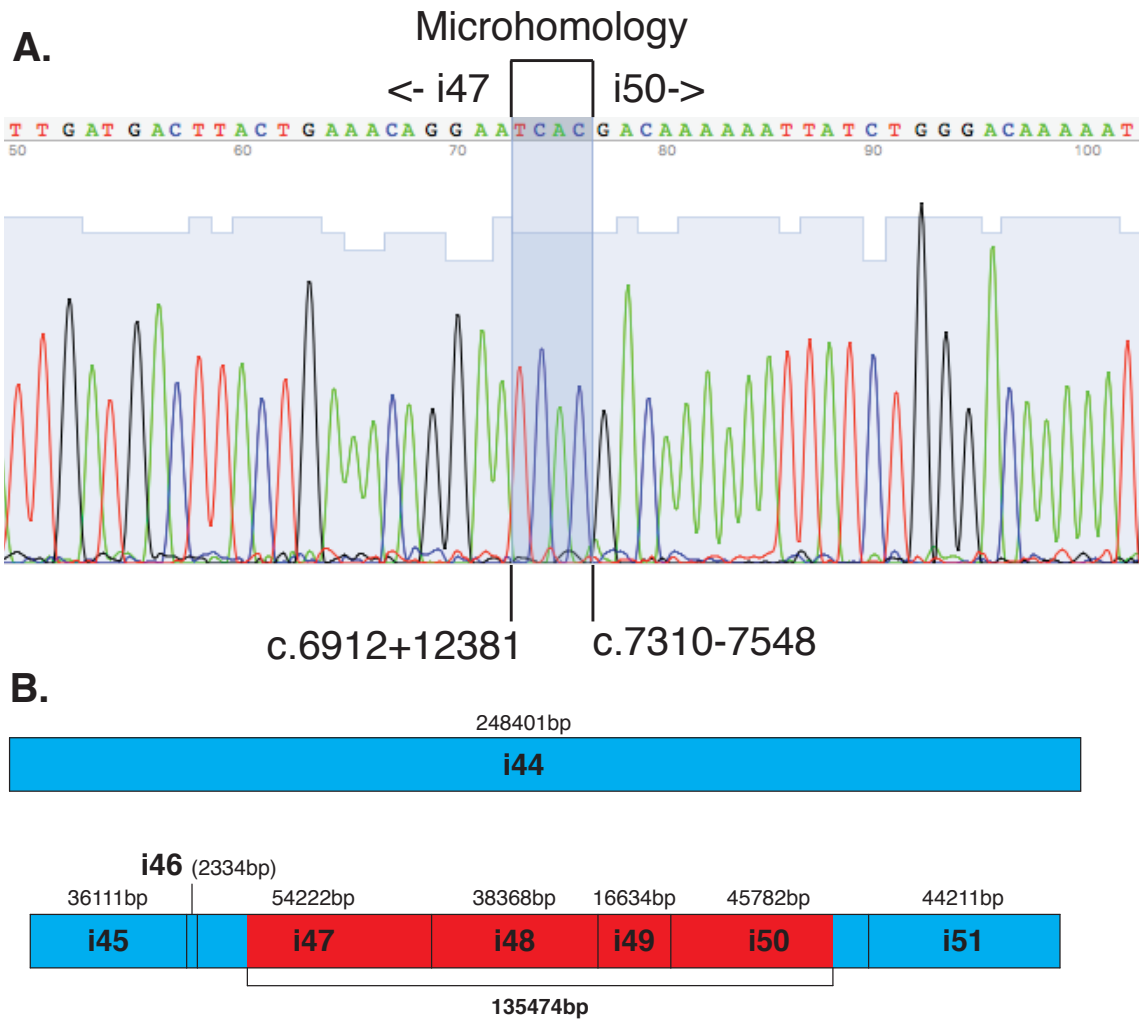


Figure 3.15: DMD gDNA exonic deletion breakpoint sequencing for cell strain 2033 (reported del. e48-50). (A) A 4bp microhomology is shown at the deletion breakpoint junction. (B) The intronic span of the deletion is highlighted in red. Intron and deletion sizes are also indicated.

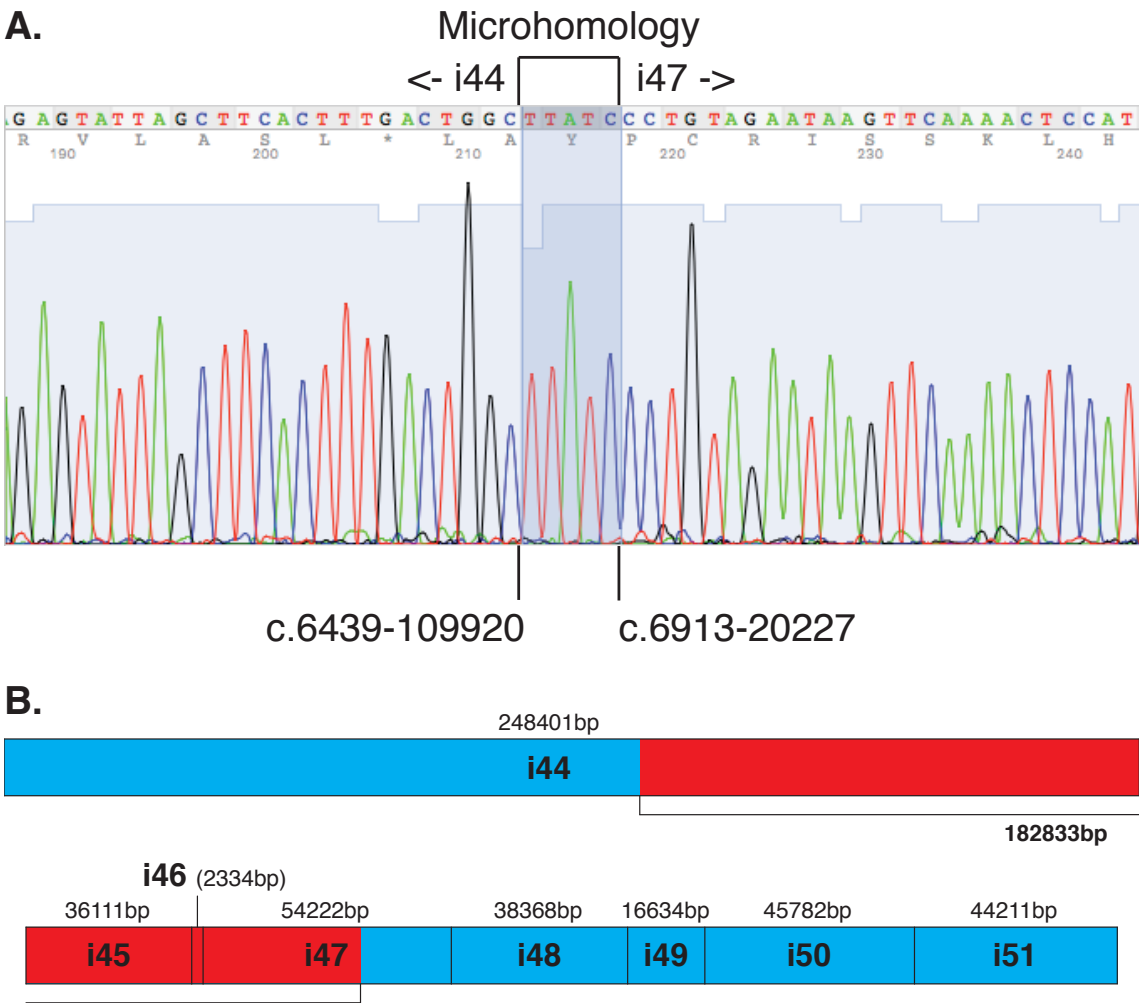


Figure 3.16: DMD gDNA exonic deletion breakpoint sequencing for cell strain 2054 (reported del. e45-47). (A) A 5bp microhomology is shown at the deletion breakpoint junction. (B) The intronic span of the deletion is highlighted in red. Intron and deletion sizes are also indicated.

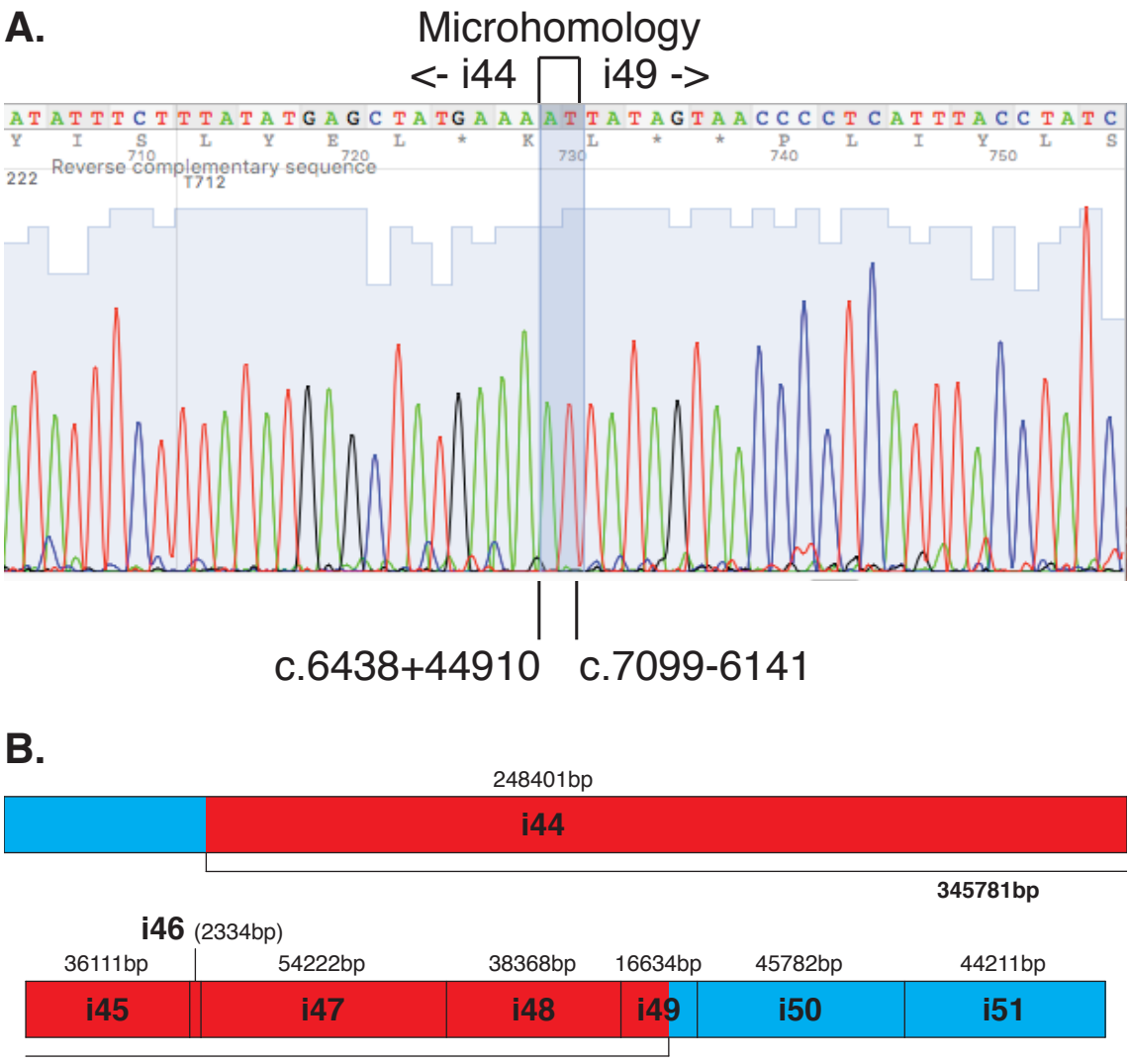


Figure 3.17: DMD gDNA exonic deletion breakpoint sequencing for cell strain 2055 (reported del. e45-49). (A) A 2bp microhomology is shown at the deletion breakpoint junction. (B) The intronic span of the deletion is highlighted in red. Intron and deletion sizes are also indicated.

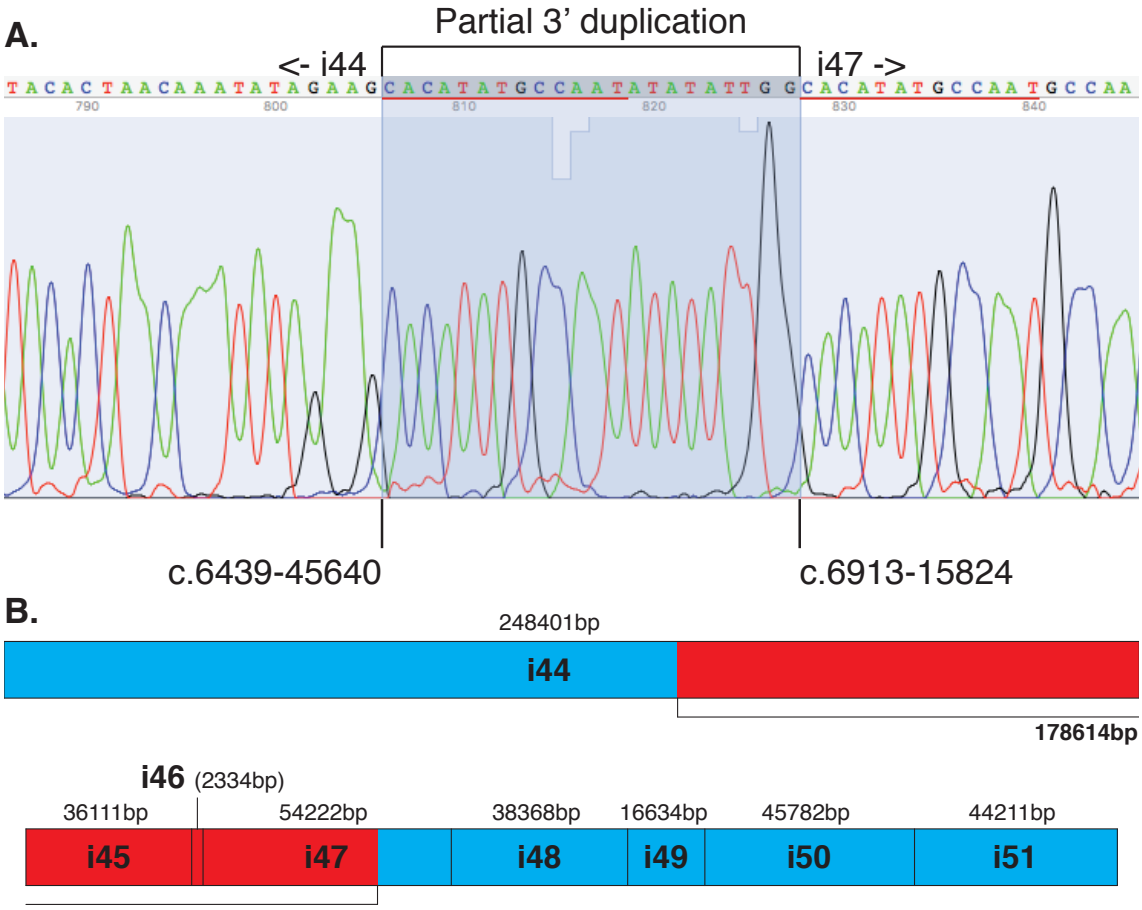


Figure 3.18: DMD gDNA exonic deletion breakpoint sequencing for cell strain 2057 (reported del. e45-47). (A) This DNA showed a 13bp duplication (underlined in red), followed by 9bp of apparently novel sequence, at the junction, creating a 22bp insert. (B) The intronic span of the deletion is highlighted in red. Intron and deletion sizes are also indicated.

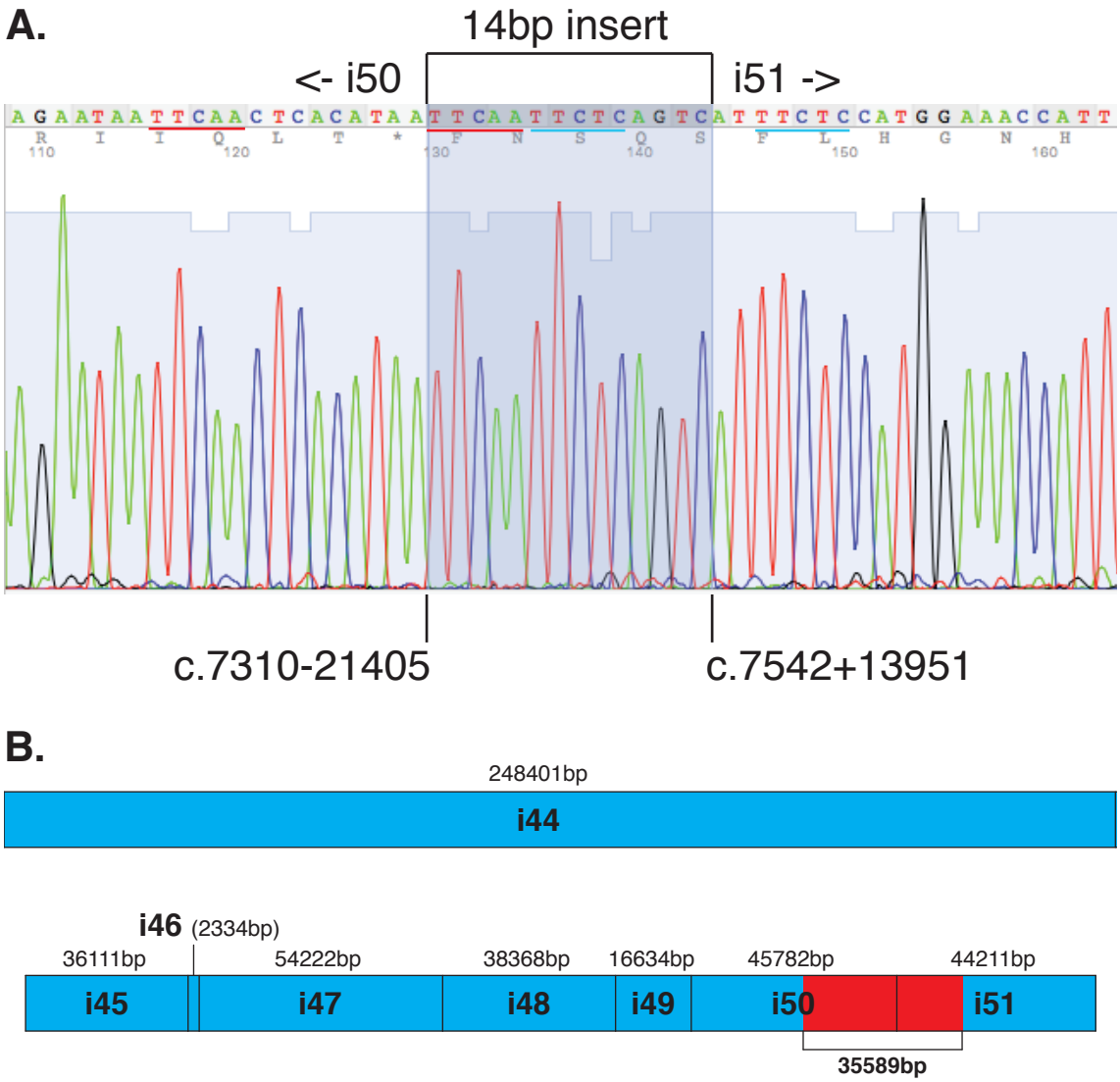


Figure 3.19: DMD gDNA exonic deletion breakpoint sequencing for cell strain 2942 (reported del. e51). (A) This DNA showed two duplications of 5bp each (underlined in red and blue) followed by 4bp of novel sequence at its junction, creating a 14bp insert. (B) The intronic span of the deletion is highlighted in red. Intron and deletion sizes are also indicated.

3.4 Discussion

Sequencing of dystrophinopathy patient transcripts to confirm whole exon deletions was performed to verify that there was a direct correspondence between the exons present in the DNA and the exon complement of the transcripts. However, this stage of the study also provided valuable context for some of the more surprising findings from the DNA sequencing.

Strain 2037 was derived from a patient diagnosed with an exon 48-49 deletion, but both exons were present in the RNA (Fig. 3.5). This study's multiplex probing of the DNA later revealed no obvious deletions in that region of the *DMD* gene (Fig. 3.11a). The only genetic variance detected in strain 2037 was the SNP close to the end of exon 48, causing a glutamine-to-lysine substitution (c.7096C>A, p.Gln2366Lys) and reducing the donor splice score of exon 48 from 89.73 to 88.77. To date, this variant has been reported 55 times in the Leiden Online Variation Database (LOVD) (Aartsma-Rus *et al.* 2006) and is classified as "Benign" in the National Centre for Biotechnology Information (NCBI) ClinVar archive (http://www.ncbi.nlm.nih.gov/clinvar/). There is therefore no reason to assume this SNP is deleterious to patient 2037; the fact that it was discovered within the region inaccurately designated as deleted is likely to be a coincidence. Together, these results indicate that either this cell strain was tagged with incorrect information, or the patient it was taken from was misdiagnosed. Though not useful to the deletion breakpoint analysis, the work on this strain exemplifies the importance of re-confirming the provided diagnoses of received cells and biopsies.

Patient 1148 was reported as having a deletion of exons 47-51. This is a deletion that is expected to preserve the open reading frame, and therefore produce a BMD phenotype. However, sequencing of the cDNA amplicon revealed that exon 46 was also missing from the mRNA transcripts (Fig. 3.2), which would disrupt the reading frame and be consistent with a diagnosis of DMD. This in itself was not proof that exon 46 was absent from the DNA; it is plausible that exon 46 was present at the genomic level but was excluded from the mature RNA during processing as an indirect result of the mutation downstream. However, DNA sequencing revealed that not only was exon 46

absent from the DNA, but so was almost the entirety of intron 45 (Fig. 3.13). This suggests that the original diagnosis provided to the patient may have been based on flawed data.

DMD exonic deletion breakpoints were successfully defined for seven cell strains. No pattern was observed in the locations of the deletion breakpoints, though in three cases the structure of the deletion junctions did exhibit consistent microhomology (Figs. 3.15, 3.16, and 3.17). This in itself may provide a clue as to how the deletions arose: during meiosis, the two X chromosomes in the mother experience an unequal crossover event. A degree of consensus is required between the two chromosome alignments at the point where strand invasion occurs, so each strand is "matched up" with the other at a site where they have at least a few bases in common. This, however, may be unlikely given the degree of homology that is required for these events to occur (Metzenberg *et al.* 1991). Alternatively, these junctions may be the result of microhomology mediated end joining repairing a severe double-stranded break within the *DMD* gene (Sfeir *et al.* 2015).

The most unusual junction sequence of the seven in the study was observed at the breakpoint of patient 1800 (del. e45-50) (Fig. 3.14). Rather than the microhomology evident in five of the six other cases, the 1800 DNA had a novel insertion of 45bp that appeared to be composed of disordered and partially nested duplications of the surrounding sequence, suggesting that this mutation may have arisen through a different mechanism than the others. The novel sequence tract could have been generated during DNA repair, following a double-stranded break. The 2942 (del. e51) DNA shows a similar feature at its junction site (Fig. 3.19), though in this case the tract of novel sequence is only 14bp long.

The DNA of patient 2057 (del. e45-47) (Fig. 3.18) also had an unusual deletion junction, with a 13bp duplication of the reference intron sequence immediately 3' of the junction followed by 9bp of novel sequence. As found in the DNA of patient 1800 (del. e45-50) (Fig. 3.14), there was no microhomology between the two intronic regions in this case, suggesting that the different forms of *DMD* deletion junctions probably arise via distinct mechanisms.

It has been suggested that exonic deletion breakpoints do not occur randomly, but instead arise as a result of local features in the genome (Verdin *et al.* 2013). Indeed, there are a number of ways that asymmetrical chromosomal rearrangements can occur, producing deletions in tandem with other rearrangements such as inversions or duplications (Stankiewicz *et al.* 2002). While it may seem that a deletion, by its very nature, would offer no clue as to how it came about, the mutations we have described here reveal that the causative mechanisms do leave behind traces of themselves at the junction site.

Breakpoint junction microhomologies such as those observed in three of the cell strains included in this study are an established phenomenon. Verdin *et al.* (2013) described 22 such microhomologies in the *FOX2* genes of separate patients, ranging in size from 1bp to 66bp. They found that microhomologies occurred at a much higher rate than would be expected if the upstream and downstream breakpoints were completely random ($p = 2.28 \times 10^{-8}$), and also that the regions around the breakpoints tend to be significantly enriched with repetitive elements.

A suggested cause of these exonic deletions is non-B DNA conformation, which can take many forms (Bacolla and Wells 2004; Inagaki *et al.* 2009). The authors posit that non-B DNA structures arising independently at each breakpoint, coupled with microhomologies at the same sites, may cause a deletion to occur as a result of errant initiation of the double-stranded break repair mechanism. It could be speculated that different combinations of non-B structures flanking a given junction might correspond to the different sequence artefacts that remain. This indicates that, with further study, it may be possible to develop a predictive model for exonic deletions that could be applied to any vertebrate genome. Once more is known about the favourabilities of the various non-B conformations to deletion initiation, the effect of microhomology on the process, etc., it would be relatively simple to scan whole genomes for potential start and end sites. A predictive map of potential breakpoint hotspots could greatly expedite future PCR-based efforts to sequence breakpoint junctions in other

exonic deletion patients, as such a strategy would avoid the need to indiscriminately scan across the entirety of large introns.

Whole Genome Sequencing (WGS) and array-comparative genomic hybridisation (array-CGH) are probably the most common methods in use today for mapping large mutations. Both offer several obvious advantages over the approach employed here: They do not require prior knowledge of the nature of the mutation; they work equally well for other large mutation types such as inversions and duplications; and in the case of Whole Genome Sequencing, once the data is generated it can be analysed further to discover other anomalies or patterns. These factors, coupled with the falling cost of Whole Genome Sequencing, indicate that it will soon become the accepted standard for large mutation discovery and investigation.

Nevertheless, there remains a strong case to be argued in favour of the PCR-based method employed in this study: Primarily, the economy of scale it offers. While, as stated above, Whole Genome Sequencing is becoming more cost effective, the typical cost per genome is still somewhat greater than US$1000 (National Human Genome Research Institute 2016), and the price-point for commercial array-CGH is similar (Emory Genetics Laboratory 2016). For many individuals, this price is prohibitively high. For researchers who might wish to investigate tens or even hundreds of mutations, Whole Genome Sequencing or array-CGH could consume a substantial amount of their operating budget, not to mention the further costs of labour needed to properly analyse the returned data. By comparison, Fractal PCR coupled with Sanger sequencing has a marginal materials cost equivalent to performing fewer than 20 PCRs per sample over the course of five or fewer thermocycler runs; and unlike the alternatives, even greater economies can be achieved if multiple genomes are investigated simultaneously. Labour costs would impact this economy variably depending on the role of the person(s) performing the work, but the techniques and expertise required are relatively basic and might make an attractive project for a graduate student.

Chapter 4

The Detection of Rare Alternative Transcripts in Normal Human RNA

4.1 Introduction

Transcript diversity is an under-explored area of research in human genetic disease, especially with regards to large genes with many exons and/or large introns, which necessarily have more scope for variation. The vast majority of human genes are alternatively spliced (Pan *et al.* 2008), and the regulation of all splicing events is highly complex and, to varying degrees, error prone (Skandelis 2016). These limitations can be magnified by mutations that affect splice sites, either altering existing sites or creating new ones. Such mutations can cause extended (Fletcher *et al.* 2013), truncated or completely skipped exons, or may lead to the emergence of pseudoexons (PEs) in the predominant transcript. However, it is plausible that some of these erroneous splicing events already exist at low levels in the transcript population and are only upregulated by the mutations that lead to their discovery.

This study used RT-PCR to investigate the possibility of PE-bearing transcripts existing in normal human RNA. Since PEs consist of intronic sequence not typically found in processed transcripts, it was expected that using them as forward primer targets for nested PCRs would provide the necessary specificity to avoid amplifying from the majority of correctly spliced transcripts. Pairing these forward primers with reverse primers in downstream canonical exons would likewise allow the ruling out of genomic DNA or post-splicing intron fragments as the template source for any amplicons produced, as inclusion of the intervening introns would make the product too massive to amplify.

The *DMD* gene was a logical choice for this project for several reasons. A large number of PEs have already been discovered in DMD patients (see Appendix B, Table B.4.1), and it may be assumed that this is due to a combination of the gene's length (+2.48Mb), its large exon count (79 for the longest and most common isoform, *Dp427m* - NCBI Reference: NG_012232.1), and the sizes of its introns, many of which are orders of magnitude larger than the human average (Sakharkar *et al.* 2004). However, the unique characteristics of the DMD gene also make discoveries about its splicing difficult to extrapolate to other genes. For this reason it was decided to include a second gene. *NF1,* like *DMD*, has a high exon count (60 in the largest isoform, - NCBI Reference: NM_001042492.2) and multiple known PEs (see Appendix B, Table B.4.2) and

is also a known disease gene - loss-of-function mutations to *NF1* cause neurofibromatosis type 1. Additionally, the NF1 gene is highly prone to splice-altering mutations in general (Messiaen et al. 2000). *NF1* is much smaller (~370Kb) than *DMD,* however, and the associated disease is autosomal dominant. All these factors made it an excellent choice as the second target gene for this study.

Using normal human RNA as a template, this study attempted to create amplicons from 21 known pseudoexon donor splice sites (PEDSSs) across two genes, *DMD* and *NF1*. If successful, this would provide an indicator of splicing fidelity and variety in these genes, and these conclusions could reasonably be extrapolated to other large, alternately spliced genes.

4.2 Specific Materials and Methods

4.2.1 Assembly of Reference Sequences

A comprehensive search of the literature was performed to discover as many *DMD* and *NF1* PEs as possible. The total set of PEs discovered was narrowed to include only those that arose as a result of single nucleotide mutations, under the assumption that these cases had the highest similarity to the reference sequence and thus were more likely to be expressed at low levels in normal individuals. PEs included in this study also needed to be large enough to comfortably accommodate two forward primers. To economise resources it was decided to only attempt PE-forward/canonical-reverse amplicons, though a canonical-forward/PE-reverse experimental design would have been equally valid.

4.2.2 RNA Extraction

RNA was extracted as per protocol from myoblast cell cultures of two different normal human cell strains.

4.2.3 Primer Design and Nested PCRs

A pair of nested forward primers were designed within each PE at 40nM scale and “Sequencing/PCR” level purity (Geneworks - http://

www.geneworks.com.au/). Each of these was paired with existing reverse primers in canonical exons downstream. RNA extracts from two normal human myoblast lines were used as the template for nested RT-PCR, which was performed as per the standard protocol (see General Materials and Methods).

4.2.4 Band Excision and Bandstabbing

The secondary PCR products were visualised on a 2% agarose gel after staining with GelRed. In cases where multiple products were generated by a single reaction, individual amplicons were isolated via the bandstab technique (Wilton *et al.* 1997). Some of the bands produced were judged to be insufficiently vivid or of the wrong size to be on-target products and were ignored. In cases where the band to be isolated was among the larger of multiple products, it was rarely possible to completely purify it of the smaller amplicons, as a small amount of these always trails behind when the gel is run. In such cases, these larger bands were bandstabbed once, excised from the post-bandstab gel, and purified on Clontech NucleoSpin Gel and PCR Clean-Up Kit columns, according to kit instructions. Fully isolated secondary/bandstab PCR products were purified on Diffinity RapidTips according to kit instructions. Purified amplicons were then submitted to AGRF (http://www.agrf.org.au) for sequencing according to AGRF's requirements.

4.3 Results

4.3.1 Summary of Investigated Pseudoexons

A thorough search of the literature discovered 11 suitable *DMD* pseudoexon mutations and 10 suitable *NF1* mutations (Tables 4.1a and 4.1b respectively - see section 4.2.1 for selection criteria). Of the 11 *DMD* mutations that initiated PE inclusion in the probands, seven were point mutations that improved the acceptor or donor splice score for the PE. Of the remaining four PEs, one (DMD pe18a) arose from a mutation that diminished the acceptor splice score for a downstream canonical exon, two (DMDpe11a, DMDpe56a) arose from local deletions that did not directly affect splice sites, and one (DMDpe77a) had no known origin.

Intron	ID	PE Size (bp)	Proband Mutation	Acceptor Splice Score	Donor Splice Score	Trans-lational Effect	Reference
i1	DMDpe1a	149	c.31+36947 **G>A**	**86.62** (70.08)	**72.56**	Frame-shift	Beroud *et al.* 2004
i11	DMDpe11a	157	c.1336+1337 **del**	**78.54**	**87.96**	Frame-shift	Malueka *et al.* 2012
i18	DMDpe18a	132	c.2622+1 **G>A**	**67.94**	**67.83**	Premature stop codon	Zhang *et al.* 2007
i25	DMDpe25a	95	c.3432+2036 **A>G**	**94.48** (77.94)	**70.66**	Frame-shift	Tuffery-Giraud *et al.* 2003
i25	DMDpe25b	203	c.3432+2240 **A>G**	**81.75**	**87.84** (70.72)	Frame-shift	Ikezawa *et al.* 1998
i26	DMDpe26a	80	c.3603+2053 **G>C**	**84.83**	**77.74**	Frame-shift	Trabelsi *et al.* 2014
i45	DMDpe45a	137	c.6614+3310 **G>T**	**72.77**	**88.44** (71.32)	Frame-shift	Gurvich *et al.* 2007
i56	DMDpe56a	166	**Flanking Deletions**	**92.52**	**87.84**	Frame-shift	Khelifi *et al.* 2011
i60	DMDpe60a	89	c.9086-15519 **G>T**	**88.11**	**93.91** (76.78)	Frame-shift	Beroud *et al.* 2004
i65	DMDpe65a	147	c.9563+1215 **A>G**	**95.33**	**94.49** (77.37)	Premature stop codon	Deburgrave *et al.* 2007
i77	DMDpe77a	151	**Unknown**	**75.39**	**86.83**	Frame-shift	Zhang *et al.* 2007

Table 4.1a: Details of 11 *DMD* PEs used as nested forward primer target sites. Reference sequence splice scores of mutation sites are shown in parentheses where appropriate.

Intron	ID	PE Size (bp)	Proband Mutation	Acceptor Splice Score	Donor Splice Score	Trans-lational Effect	Reference
i3	NF1pe3a	108	c.288+2025 **T>G**	**90.06**	**83.43** (66.31)	Premature Stop Codon	Pros *et al.* 2009
i6	NF1pe6a	132	c.889-942 **G>T**	**72.71** (66.24)	**81.82**	Premature Stop Codon	Pros *et al.* 2008
i6	NF1pe6b	58	c.888+651 **T>A**	**88.37** (71.82)	**78.75**	Frame-shift	Messiaen and Wimmer 2008 *in* Kannu 2013
i10a	NF1pe10a1	76	c.1393-592 **A>G**	**81.29**	**79.54** (67.69)	Frame-shift	Pros *et al.* 2008
i10b	NF1pe10b1	54	c.1527+1159 **C>T**	**86.08**	**82.91** (65.79)	Premature Stop Codon	Pros *et al.* 2008
i19a	NF1pe19a1	99	c.3198-314 **G>A**	**84.57** (68.02)	**87.84**	Premature Stop Codon	Fernandez-Rodriguez *et al.* 2011
i23a	NF1pe23a1	80	c.4173+278 **A >G**	**87.04** (70.50)	**82.55**	Frame-shift	Kannu *et al.* 2013
i30	NF1pe30a	177	c.5749+332 **A>G**	**75.38**	**86.77** (69.65)	Premature Stop Codon	Perrin *et al.* 1996
i30	NF1pe30b	171	c.5750-279 **A>G**	**84.01** (78.63)	**78.63**	Premature Stop Codon	Raponi *et al.* 1996

Table 4.1b: Details of 9 *NF1* PEs used as nested forward primer target sites. Reference sequence splice scores of mutation sites are shown in parentheses where appropriate.

4.3.2 DMD *Pseudoexon PCRs*

Primers were designed or selected from pre-existing stock as needed. Details are listed in Table 4.3. Nested RT-PCRs were performed as per Materials and Methods sections 2.1 and 2.2. A total of five on-target amplicons (Figs. 4.1a and 4.1b) were successfully isolated and sequenced (Figs. 4.2a - 4.2d) from the PCR products.

PE ID	Outer Forward	Outer Reverse	Inner Forward	Inner Reverse	Amplicon	Expected Size (bp)
DMDpe1a	6043	5308	6044	4952	pe1a -> e10	1020
DMDpe11a	6045	4332	6046	4334	pe11a -> e18	921
DMDpe18a	6047	4344	6048	4336	pe18a -> e25	1137
DMDpe25a	6049	4620	6050	4621	pe25a -> e33	1218
DMDpe25b	6051	4620	6052	4621	pe25b -> e33	1244
DMDpe26a	6053	4620	6054	4621	pe26a -> e33	1021
DMDpe45a	6055	5306	6056	4851	pe45a -> e51	779
DMDpe56a	6057	4358	6058	4360	pe56a -> e58	511
DMDpe60a	6059	4328	6060	4628	pe60a -> e67	729
DMDpe65a	6061	4891	6062	4889	pe65a -> e77	1370
DMDpe77a	6063	4206	6064	4206	pe77a -> e79	525
+ve	4353	5306	4456	5464	e44 -> e51	934
-v1	None	None	None	None	None	0
-v2	None	None	None	None	None	0

Table 4.2: Nested PCRs of normal human *DMD* transcripts, showing amplicon spans and expected sizes. Primer sequences are listed in Appendix D.

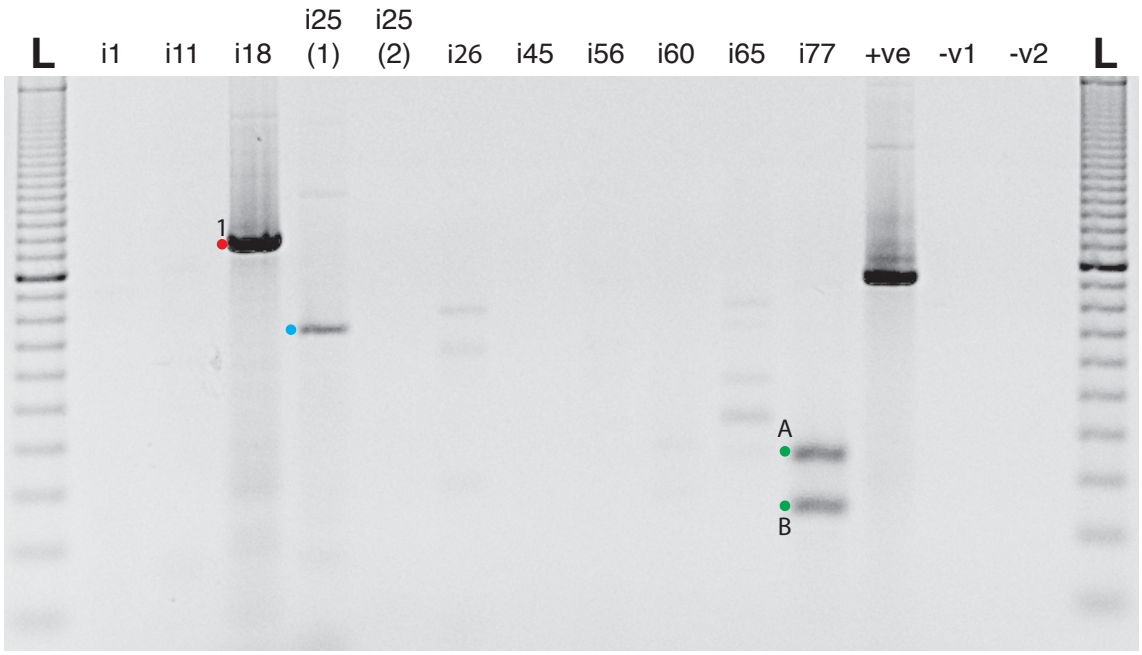


Figure 4.1a: PCR products of searches for PE-bearing *DMD* transcripts in normal human myoblast RNA (Sample 1). "L" indicates 100bp ladder size standard. One distinct amplicon (1, indicated with red dot) was successfully isolated and sequenced. Two more (A and B, indicated with green dots) were identified as off-target products. One additional band (blue dot) was isolated but could not be confirmed as a distinct, on target transcript. Results for the on-target amplicon are listed by number in Table 4.4.

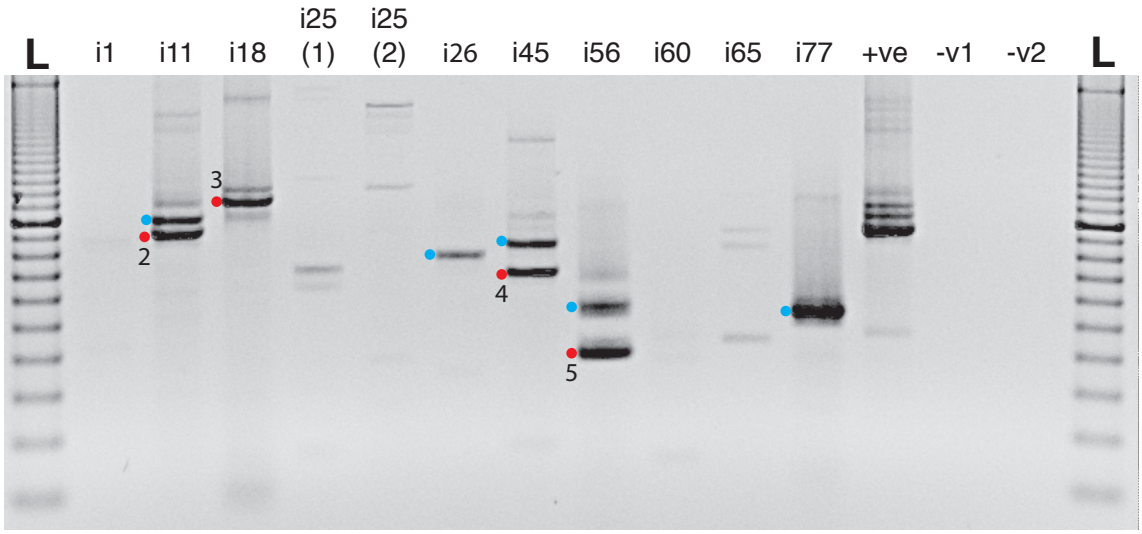


Figure 4.1b: PCR products of searches for PE-bearing *DMD* transcripts in normal human myoblast RNA (Sample 2). "L" indicates 100bp ladder size standard. Four distinct amplicons (2-5, indicated with red dots) were successfully isolated and sequenced. Five additional bands (blue dots) were isolated but could not be confirmed as distinct, on target transcripts. Results for each amplicon are listed by number in Table 4.4.

4.3.3 NF1 *pseudoexon PCRs*

Primers were designed or selected from pre-existing stock as needed. Details are listed in Table 4.3. Nested RT-PCRs were performed as per Materials and Methods sections 2.1 and 2.2. A total of nine on-target amplicons (Figs. 4.1c and 4.1d) were successfully isolated and sequenced (Figs. 4.2e - 4.2j) from the PCR products.

PE ID	Outer Forward	Outer Reverse	Inner Forward	Inner Reverse	Amplicon	Expected Size (bp)
NF1pe3a	6187	6207	6188	6208	pe3a -> e8	866
NF1pe6a	6189	6207	6190	6208	pe6a -> e8	249
NF1pe6b	6191	6207	6192	6208	pe6b -> e8	236
NF1pe10a1	6193	6209	6194	6210	pe10a1 -> e13	716
NF1pe10b1	6195	6209	6196	6210	pe10b1 -> e13	572
NF1pe19a1	6197	6211	6198	6212	pe19a1 -> e25	1140
NF1pe23a1	6199	6211	6200	6212	pe23a1 -> e25	239
NF1pe30a	6201	6213	6202	6214	pe30a -> e33	544
NF1pe30b	6203	6215	6204	6214	pe30b -> e33	528
+ve	6397	6213	6398	6214	e29 -> e33	722
-v1	None	None	None	None	None	0
-v2	None	None	None	None	None	0

Table 4.3: Nested PCRs of normal human *NF1* transcripts, showing amplicon spans and expected sizes. Results for each amplicon are listed by number in Table 4. Primer sequences are listed in Appendix D.

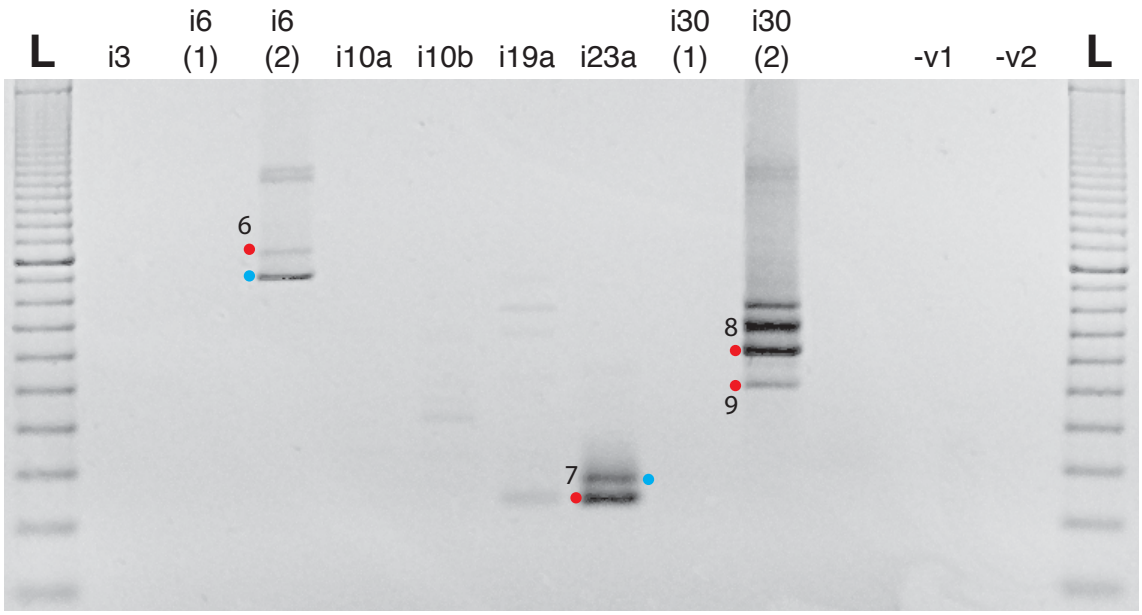


Figure 4.1c: PCR products of searches for PE-bearing *NF1* transcripts in normal human myoblast RNA (Sample 1). "L" indicates 100bp ladder size standard. Four distinct amplicons (6-9, indicated with red dots) were successfully isolated and sequenced. Two additional bands (blue dots) were isolated but could not be confirmed as distinct, on target transcripts. Results for each amplicon are listed by number in Table 4.4.

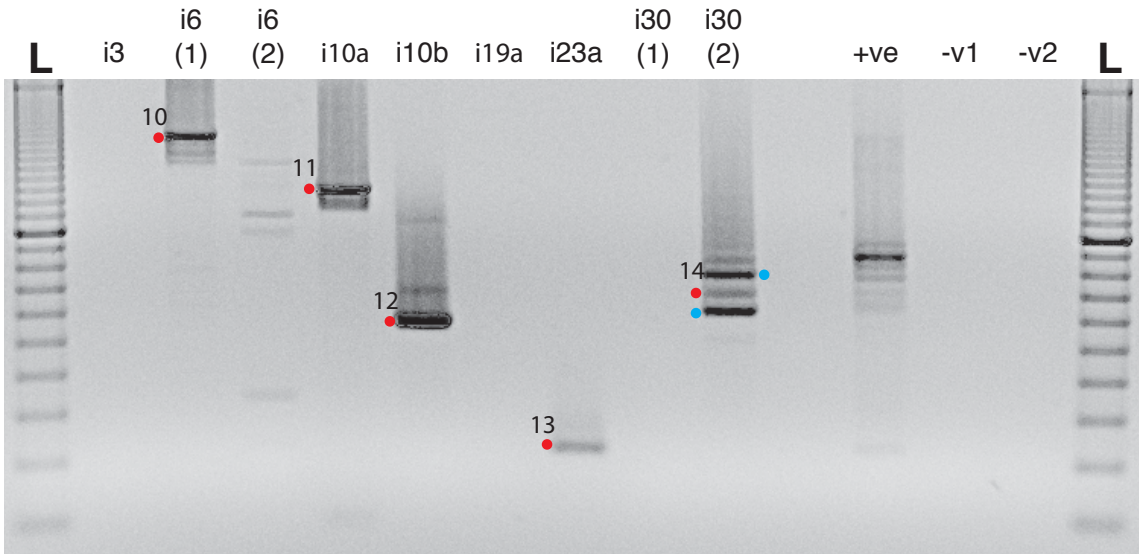


Figure 4.1d: PCR products of searches for PE-bearing *NF1* transcripts in normal human myoblast RNA (Sample 2). "L" indicates 100bp ladder size standard. Five distinct amplicons (10-14, indicated with red dots) were successfully isolated and sequenced. Two additional bands (blue dot) were isolated but could not be confirmed as distinct, on target transcripts. Results for each amplicon are listed by number in Table 4.4.

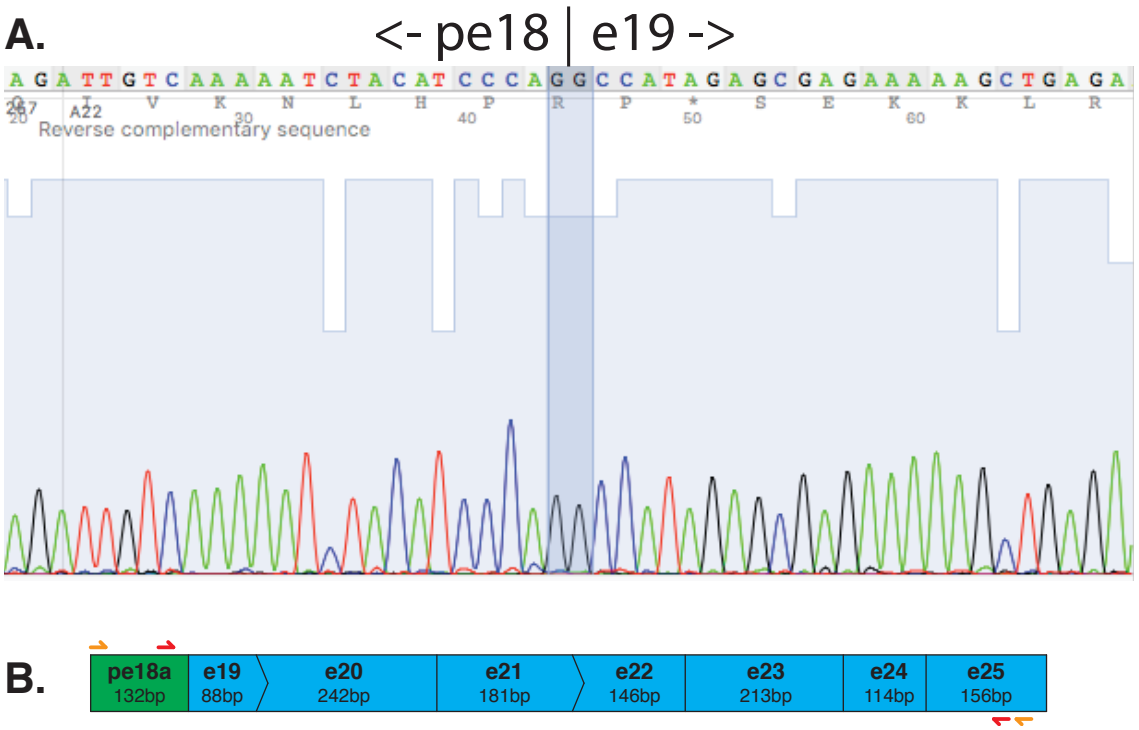


Figure 4.2a: Sequencing of putative pe18a-bearing *DMD* transcript from normal human myoblast RNA (sample 1). (A) This sequence confirmed the presence of transcripts bearing the pe18a donor splice site. An identical result was obtained from sample 2 (not shown). (B) Diagram of the amplified region. Orange and red arrows indicate positions of outer and inner primers respectively. Non-canonically spliced regions are coloured green, canonical exons blue.

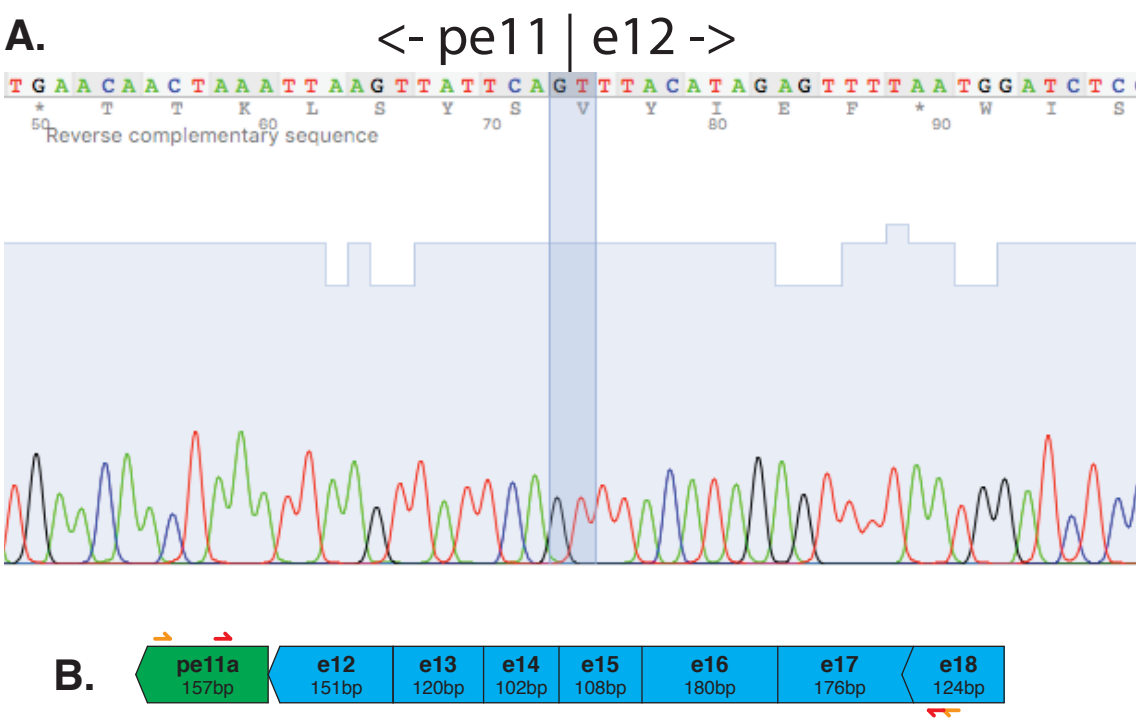


Figure 4.2b: Sequencing of putative pe11a-bearing *DMD* transcript from normal human myoblast RNA (sample 2). (A) This sequence confirmed the presence of transcripts bearing the pe11a donor splice site. (B) Diagram of the amplified region. Orange and red arrows indicate positions of outer and inner primers respectively. Non-canonically spliced regions are coloured green, canonical exons blue.

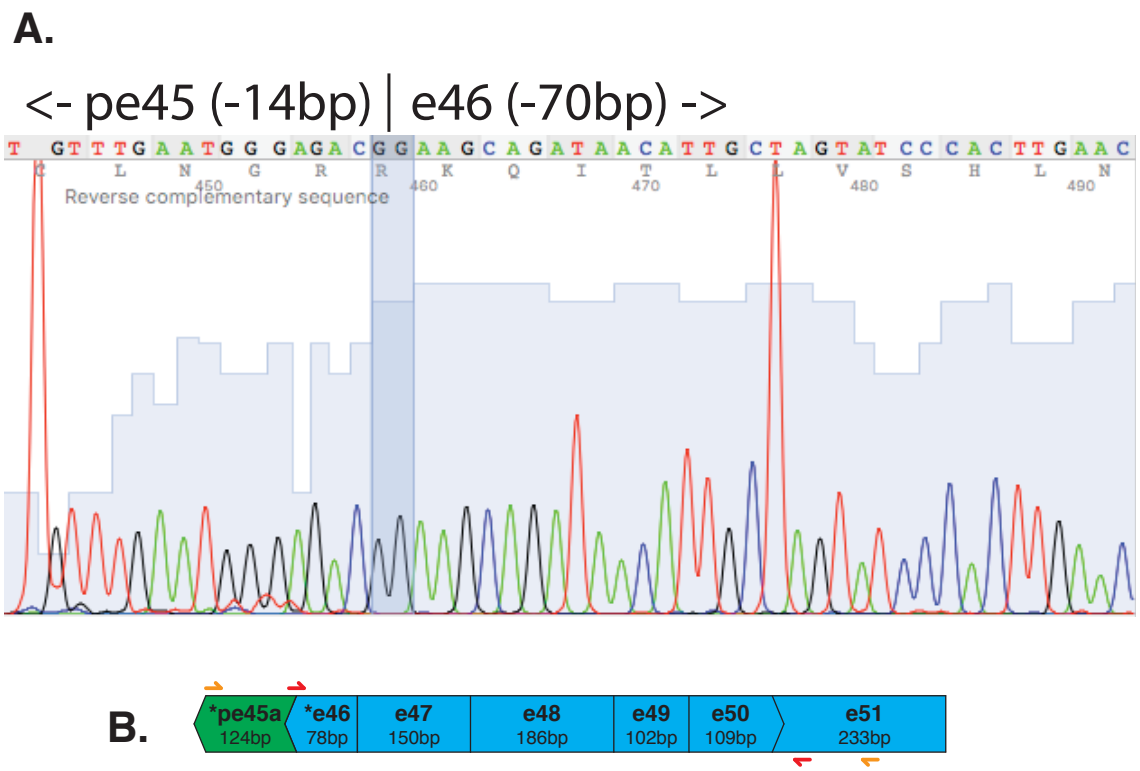


Figure 4.2c: Sequencing of putative pe45a-bearing *DMD* transcript from normal human myoblast RNA (sample 2). (A) This amplicon was on-target but revealed alternative donor and acceptor sites at the PE junction, meaning 14bp were spliced from the end of pe45 and 70bp were spliced from the beginning of exon 46. (B) Diagram of the amplified region. Orange and red arrows indicate positions of outer and inner primers respectively. Non-canonically spliced regions are coloured green, canonical exons blue.

Figure 4.2d: Sequencing of putative pe56a-bearing *DMD* transcript from normal human myoblast RNA (sample 2). (A) This sequence confirmed the presence of transcripts bearing the pe56a donor splice site. (B) Diagram of the amplified region. Orange and red arrows indicate positions of outer and inner primers respectively. Non-canonically spliced regions are coloured green, canonical exons blue.

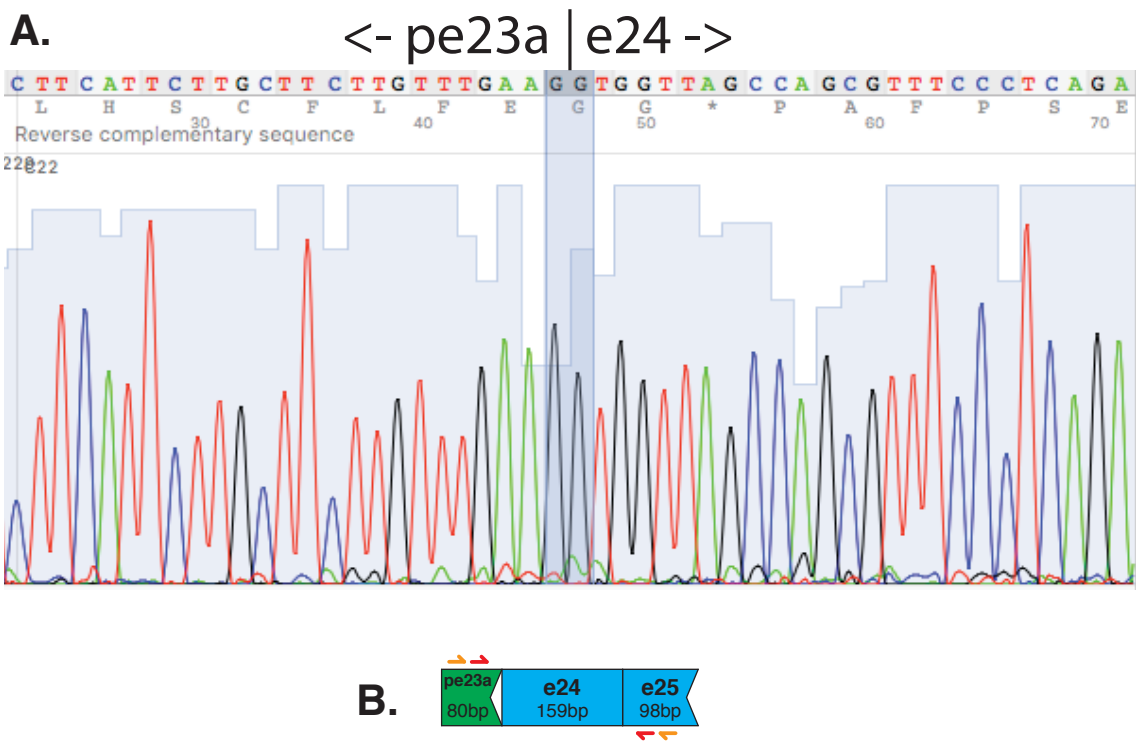


Figure 4.2e: Sequencing of putative pe23a-bearing *NF1* transcript from normal human myoblast RNA (sample 1). (A) This sequence confirmed the presence of transcripts bearing the PE donor splice site. An identical result was obtained from sample 2 (not shown). (B) Diagram of the amplified region. Orange and red arrows indicate positions of outer and inner primers respectively. Non-canonically spliced regions are coloured green, canonical exons blue.

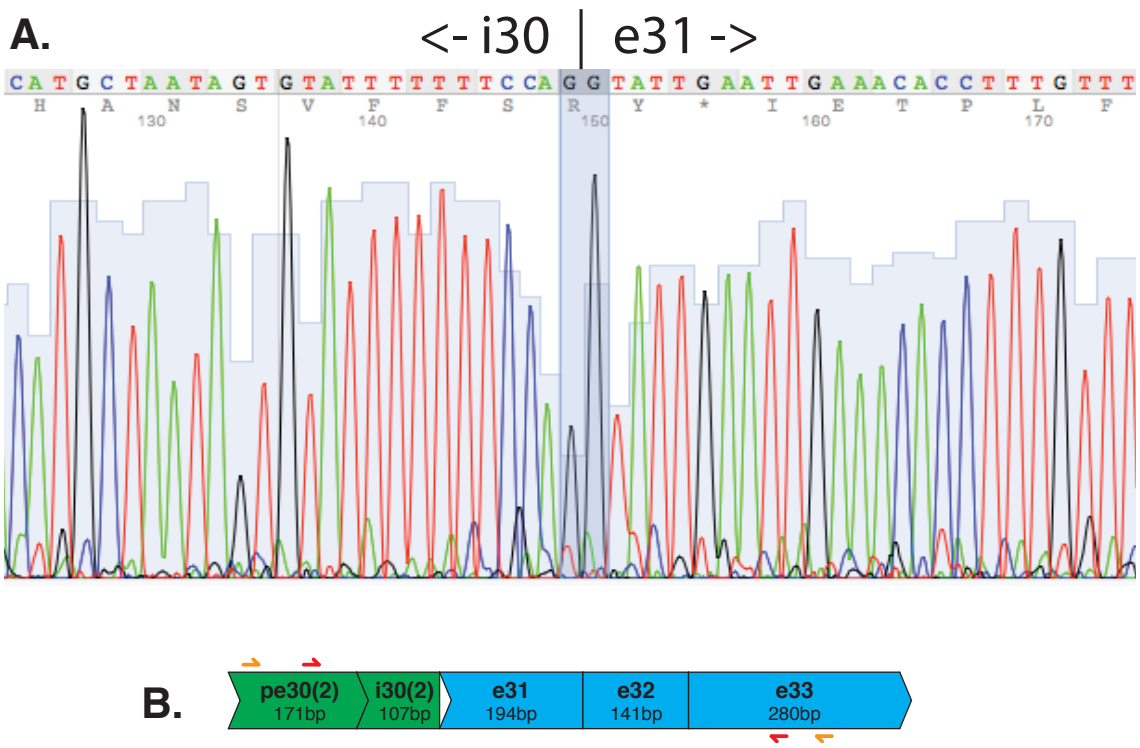


Figure 4.2f: Sequencing of putative pe30b-bearing *NF1* transcript from normal human myoblast RNA (sample 1). (A) This amplicon was larger than expected due to the inclusion of at least part of the 3' end of i30 in the transcript. An identical result was obtained from sample 2 (not shown). (B) Diagram of the amplified region. Orange and red arrows indicate positions of outer and inner primers respectively. Non-canonically spliced regions are coloured green, canonical exons blue.

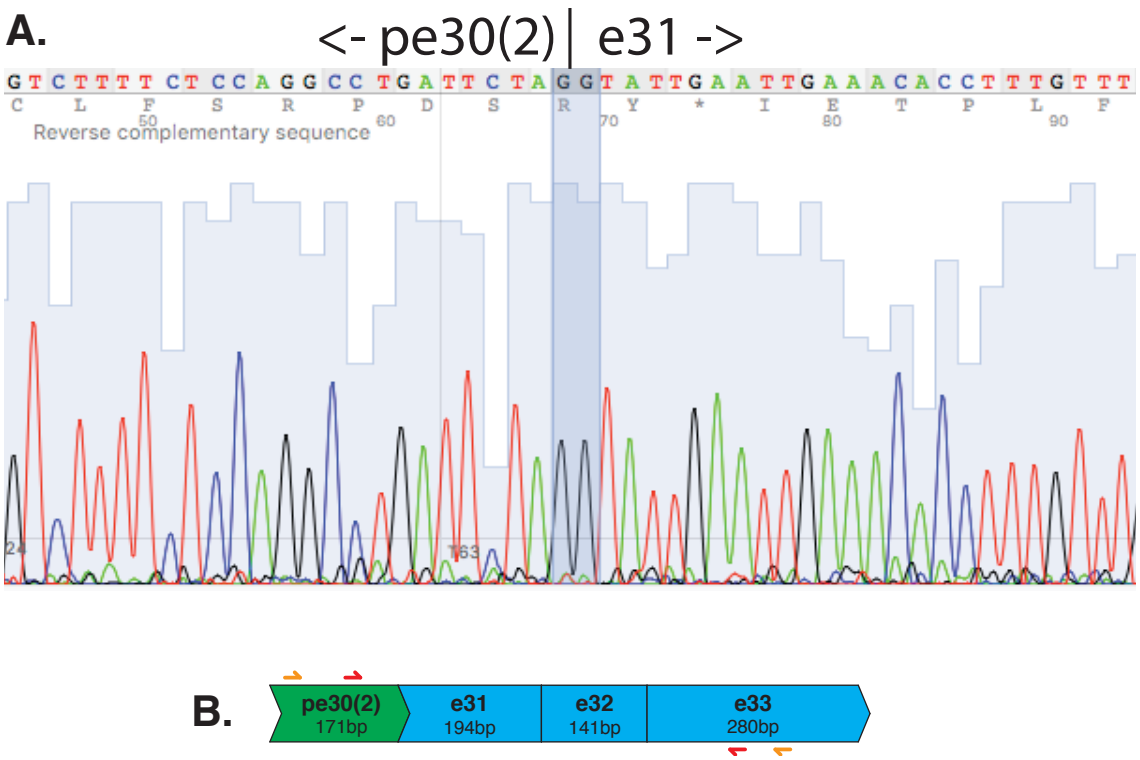


Figure 4.2g: Sequencing of putative pe30b-bearing *NF1* transcript from normal human myoblast RNA (sample 1). (A) This sequence confirmed the presence of transcripts bearing the pe30b donor splice site. (B) Diagram of the amplified region. Orange and red arrows indicate positions of outer and inner primers respectively. Non-canonically spliced regions are coloured green, canonical exons blue.

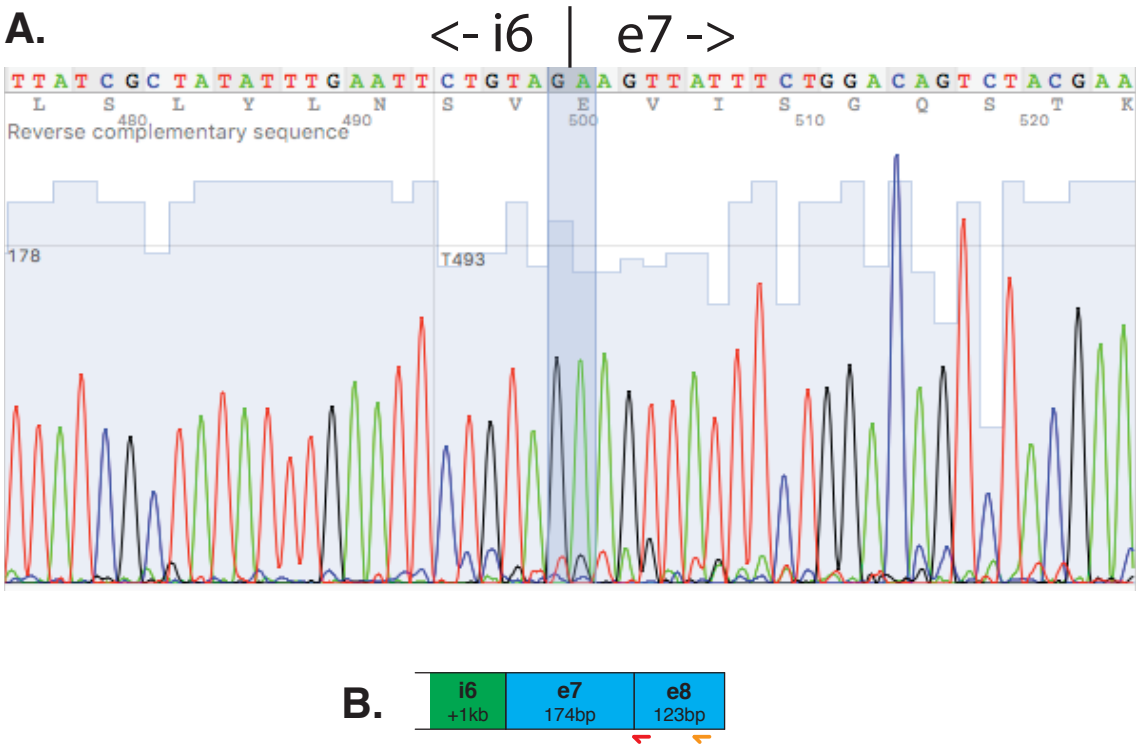


Figure 4.2h: Sequencing of putative pe6a-bearing *NF1* transcript from normal human myoblast RNA (sample 2). (A) This amplicon was larger than expected due to the inclusion of part of the 3' end of i6 in the transcript. (B) Diagram of the amplified region. Orange and red arrows indicate positions of outer and inner primers respectively. Non-canonically spliced regions are coloured green, canonical exons blue. Forward primers are not shown as the sequencing of the 5' end of the transcript could not confirm their positions.

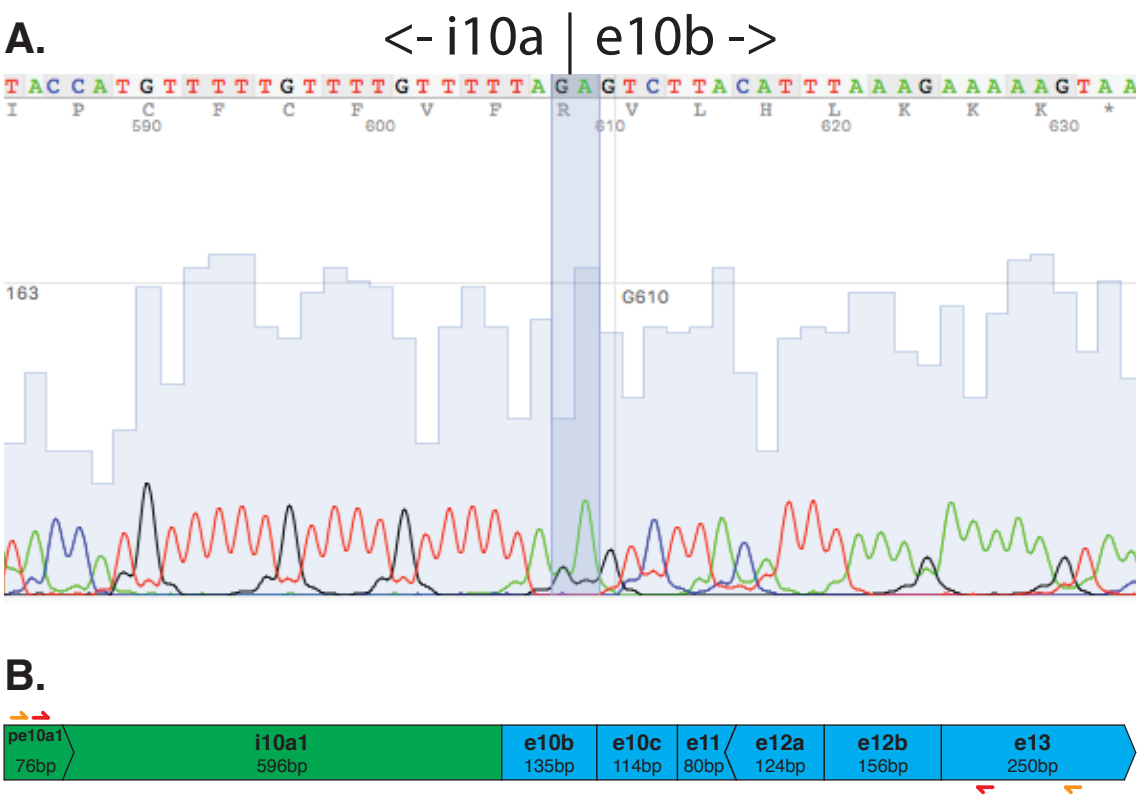


Figure 4.2i: Sequencing of putative pe10a1-bearing *NF1* transcript from normal human myoblast RNA (sample 2). (A) This amplicon was larger than expected due to the inclusion of part of the 3' end of i10a in the transcript. (B) Diagram of the amplified region. Orange and red arrows indicate positions of outer and inner primers respectively. Non-canonically spliced regions are coloured green, canonical exons blue.

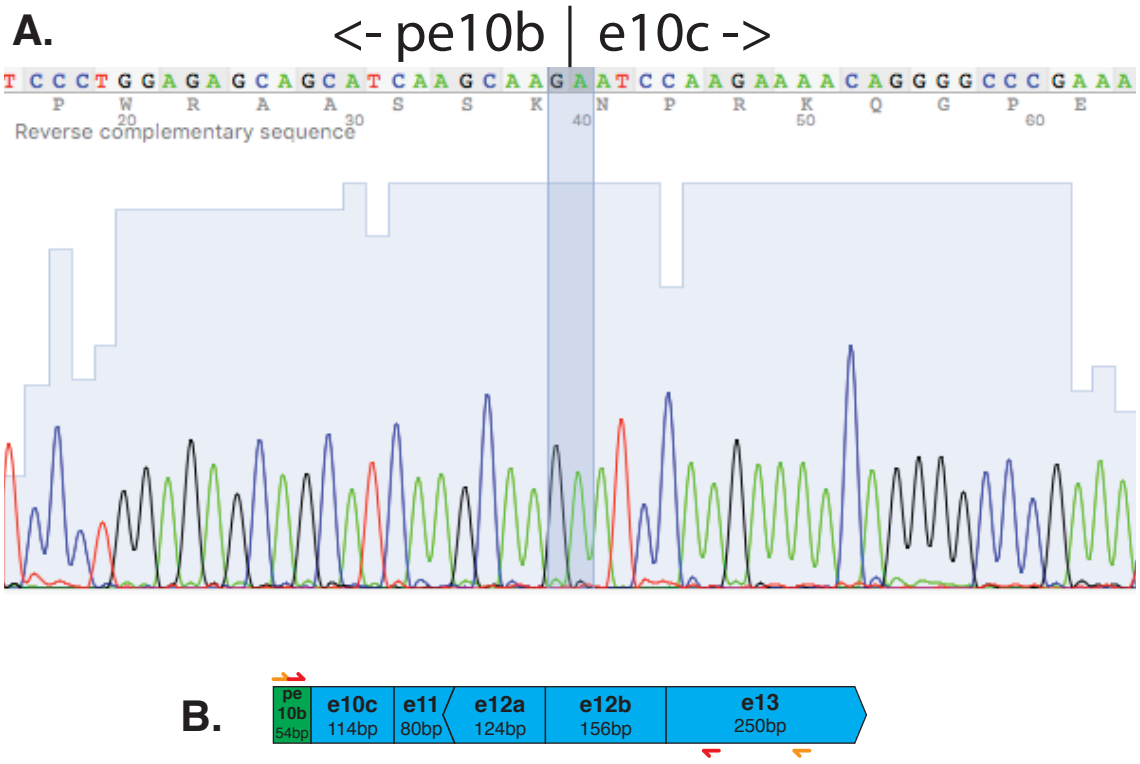


Figure 4.2j: Sequencing of putative pe10b1-bearing *NF1* transcript from normal human myoblast RNA (sample 2). (A) This sequence confirmed the presence of transcripts bearing the pe10b1 donor splice site. (B) Diagram of the amplified region. Orange and red arrows indicate positions of outer and inner primers respectively. Non-canonically spliced regions are coloured green, canonical exons blue.

4.3.4 Results summary

Results of the *DMD* and *NF1* RT-PCRs and sequencing are summarised in Table 4.4.

#	Experiment	Gel Band	Expected Size	Actual Size	Sequence Result
1	DMD, N1	i18	1137bp	Same	Confirmed PEDSS inclusion
2	DMD, N2	i11S	921bp	Same	Confirmed PEDSS inclusion
3	DMD, N2	i18	1137bp	Same	Confirmed PEDSS inclusion
4	DMD, N2	i45S	779bp	695bp	Confirmed PEDSS inclusion, alternative 5' and 3' splice sites
5	DMD, N2	i56S	390bp	Same	Confirmed PEDSS inclusion
6	NF1, N1	i6(2)L	236bp	~900bp	Confirmed partial 3' intron inclusion
7	NF1, N1	i23a	239bp	Same	Confirmed PEDSS inclusion
8	NF1, N1	i30(2)L	528bp	~700bp	Confirmed partial 3' intron inclusion
9	NF1, N1	i30(2)S	528bp	Same	Confirmed PEDSS inclusion
10	NF1, N2	i6(1)	249bp	~1900bp	Confirmed partial 3' intron inclusion
11	NF1, N2	i10a	716bp	~1250bp	Confirmed partial 3' intron inclusion
12	NF1, N2	i10b	572bp	Same	Confirmed PEDSS inclusion
13	NF1, N2	i23a	239bp	Same	Confirmed PEDSS inclusion
14	NF1, N2	i30(2)L	528bp	~800bp	Confirmed partial 3' intron inclusion

Table 4.4: Summary of results for pseudoexon donor splice site (PEDSS) searches. In eight cases (1, 2, 3, 5, 7, 9, 12 and 13), PEDSS inclusion was confirmed. In another five cases (6, 8, 10, 11 and 14), inclusion of the complete intervening sequence of intron was observed. In one case (4), PEDSS inclusion was observed but novel donor and acceptor splice sites were used, producing a smaller amplicon.

4.4 Discussion

While the techniques employed in this study provide a focused window into the scope of RNA transcript complexity, what they reveal nonetheless has implications for the understanding of splicing regulation. Though only preliminary data, these results indicate that, in some cases, PEs initiated by pathogenic point mutations are likely to already exist at low levels in the mature mRNA population - either as intermediate splicing products or as a result of imperfect spliceosome fidelity. (These are not likely to be pathogenic, given that all the PEs examined in this study produce premature stop codons or frame shifts in the transcript and do not create toxic gains-of-function.) It may be that canonical, constitutive splicing is not an all-or-nothing phenomenon, wherein the best splice site is utilised 100% of the time; rather, it appears to proceed in a probabilistic manner, with the canonical sites predominating but weaker alternative sites also being fractionally utilised. This is a phenomenon that could be further explored across multiple genes, cell lines and tissue types. Given that the *DMD* transcript is matured to such diverse isoforms across the different organs of the body, it would be expected that the patterns of its cryptic splicing would change also.

We discovered three cases where the entire 3' end of the intron following the forward primer was included in the transcript. From the existing evidence, it is impossible to say how much of the upstream intron has been included in these instances. It should be noted, however, that these intronic inclusions were only identified because those PEs are located close to the 3' end of their introns, and the longer amplicon is therefore still within the size scope of a standard nested PCR. It may be that partial or whole intronic inclusions were more prevalent, but in most cases could not be detected by this experimental design. Future studies might investigate this possibility more methodically by probing normal RNA with primers targeted close to either end of each intron in a given gene.

While the replicates of this experiment are low (n = 2), it is still notable that so many more rare transcripts were detected from the *NF1* gene than from *DMD* (9 vs. 5). This may be explained by this gene's heightened propensity for illegitimate splicing even in apparently unmutated individuals (Wimmer *et al.* 2000), though the question of why it has this propensity in the first place

remains unanswered. The total known pseudoexon counts for both genes (Tables 4.1a and 4.1b) are also unusual. *DMD* is more than six times the size of *NF1* and has greater intron content, so we might expect that there would be proportionally more PEs discovered within it; but this is not the case. While we found more PEs for DMD than for *NF1* overall, on a like-for-like comparison of the numbers of PEs initiated by point mutations, both genes are very similar. There are several factors that may account for this disparity. *NF1* undergoes a smaller variety of alternative splicing than *DMD*, with only three of its exons thus far identified to be differentially spliced: 9a/9br (Danglot *et al.* 1995), 23a (Nishi *et al.* 1991), and 48a (Marchuk *et al.* 1991). *DMD*, by comparison, has at least 19 unique isoforms (Leiden Muscular Dystrophy Pages 2006). It may be that differences in splicing regulation in alternatively spliced genes render them less susceptible to being mis-spliced (Chen and Manley 2009).

Nonsense mediated decay (NMD) likely evolved as a method for cleaning up incorrectly spliced RNA transcripts (Atkinson *et al.* 2009), such as those identified in this experiment. NMD targets transcripts that, as the result of either a frameshifting genomic mutation or splicing event, have gained a premature stop codon within the open reading frame (Kervestin *et al.* 2012). The NMD process prevents the RNA from being unnecessarily translated into functionless or harmful protein, which would at best waste metabolic resources and at worst kill the cell. This, then, raises the question of why the NMD mechanism has allowed this menagerie of atypical transcripts to exist. The simplest explanation is that there is a delay in the initiation of NMD, which gives mis-spliced mRNAs a drastically shortened but non-zero lifespan within the cell. When the source cell cultures were harvested for total RNA, a “snapshot” was taken of the transcript population, and it is perhaps unsurprising that within these snapshots were a tiny minority of aberrant transcripts that had not yet been detected and broken down by NMD. It may also be that there were some cases where the inclusion of all or part of the pseudoexon would not disrupt the reading frame, and these would be correspondingly better tolerated by NMD.

An alternative explanation for the existence of these rare transcripts is provided by Gazzoli et al. (2016). The authors have recently presented evidence that splicing for many large DMD introns proceeds in a segmented, multi-step

fashion. The authors predicted 414 potential intra-intronic splice sites in the *DMD* gene transcript. Though they were only able to empirically test 13 of the 414 sites, they succeeded in verifying eight, suggesting that there may be many more such junctions that have yet to be discovered. Interestingly, two pseudoexon donor splice sites detected in this study were exact matches to those amongst Gazzoli *et al.*'s unverified predictions (Table 4.4, rows 1, 3 and 5). This suggests that, in these two cases at least, the detected splice junctions may not have represented true pseudoexon inclusion events, but rather intermediate forms of the *DMD* transcript as it matured through multi-step splicing.

While the goal of this study was to detect rare inclusions of known PEDSSs in target gene transcripts, it also inadvertently detected cases of partial intron inclusion. Like PEs, these are most often detected when their inclusion is triggered by an activating mutation (Fletcher *et al.* 2013; Greer *et al.* 2015). Also noted was one case of a combined PEDSS inclusion with an alternative donor splice site and alternative acceptor splice sites (Fig. 4.2c). The only type of splicing disruption not detected was a naturally skipped canonical exon, probably because the primer design strategy of the experiment gives it no more advantage in detecting such transcripts, should they exist, than a conventional RT-PCR would.

In terms of providing a complete picture of transcript diversity, whole transcriptome sequencing (WTS, or RNA-Seq) is in some ways superior to the techniques demonstrated here, revealing as it does complete sequences for the majority of transcripts in a quantitative fashion. Others have noted, however, that it is limited in its ability to detect very low copy number transcripts (Ameur *et al.* 2011). For this reason, exotic transcript PCR remains a worthwhile strategy for some purposes. The simplicity of its design makes it an attractive option for researchers who may not have access to WTS technology, or who are interested in transcript variation in a single splicing region across multiple cell lines. The specificity of this strategy also allows it to detect transcripts that could be missed by WTS. This strategy could also act as a "stepping stone" to these techniques, verifying that transcript variance does exist for a given gene, before proceeding to more in-depth analyses.

Chapter 5

Evaluation of Antisense Oligomer Splicing Motif Targets

5.1 Introduction

In the past, bioinformatics has played only a limited role in the design and optimisation of exon skipping AOs (Pramono *et al.* 2012). Web tools such as Splice Aid 2 (Piva *et al.* 2012) have made it possible to easily and quickly visualise the positions of splice factor motifs in any input sequence, and this has enabled an approach to AO design that is considerably better than random. However, while it is relatively simple to target AOs to enhancer sites and avoid silencers, this approach does not always yield an effective splice modifying AO, and it does not account for which splice factors are important and to what extent.

To date, there have been very few bioinformatic analyses of exon-skipping AO targets in the *DMD* gene. Aartsma-Rus *et al.* (2009) examined a number of locational and conformational properties of exon skipping AOs and their target regions, as well as their coincidence with some serine-rich protein binding sites. The authors found that a key mechanism of action for AOs targeted within exons is to block splice enhancer motifs. Echigoya *et al.* (2015) performed a similar study and found that the binding energetics and proximity to the splice acceptor site were also important factors. However, these studies considered very few splicing factors (four and seven respectively) for correlations to AO efficacy. This leaves the great majority of known splicing factors and their motifs unexplored as potential guides to exon-skipping AO design.

The existence of a substantial body of data on antisense oligomer design, accumulated by our group, presented a unique opportunity for bioinformatic analysis to aid splice modifying oligomer design. Using the Python programming language, scripts were constructed that would compare the frequency of splicing factor motifs at “best-known” oligo target sites versus less effective sites. If relationships were discovered, these could be used to construct a predictive model for identifying exon skipping target sites. This model would be a valuable tool for targeting exon-skipping AOs to other gene transcripts, thereby economising on the resources consumed by empirical oligomer design and testing.

5.2 Materials and Methods

5.2.1 Dataset Assembly and Curation

Three datasets formed the basis of the analysis: Splicing factor motifs (the "splice factors" dataset), *DMD* exon sequences which include 50bp of flanking intron on either side of each exon (the "exons" dataset), and the sequences of the AOs together with their efficacy data (the "AOs" dataset). All datasets were formatted as .csv files before processing.

5.2.1.1 The AOs Dataset

The AOs dataset was assembled from our group's experimental records. The efficacy of each AO was simplified to a Yes/No categorical value, depending on whether or not it was the best AO out of all those tested for skipping the target exon. While a ranking system was considered, it was decided that a Yes/No system would minimise the data "noise" inherent to ranking AOs trialled across a variety of operators, locations and times. Usually there was only one "best" AO per exon, although multiple "bests" were counted for a single exon in cases where more than one non-overlapping AO proved equally effective at inducing exon skipping. AOs were further curated, using a Python script, to include only those that were fully complementary to their sequence target. A summary of the curated AO dataset is shown in Appendix C, Table C.5.1.

5.2.1.2 The Exons Dataset

The exons dataset was assembled from the *DMD* reference sequence (NG_012232.1).

5.2.1.3 The Splice Factors Dataset

The splice factors dataset was assembled from the full table of 71 splicing factors and their binding motifs, available from the Splice Aid 2 web tool (Piva *et al.* 2012). All factors and motifs were retained, though it was necessary to remove some motif duplicates that would have otherwise biased the results. A summary of the curated splice factors dataset is shown in the Appendix C, Table C.5.2.

5.2.2 Data Processing

This phase is described in Fig. 5.1. (see below). The source code for the Python script used is shown in Appendix C, section C.5.1, and descriptions of how to correctly format the input tables, and interpret the output table generated by this script, are given in Appendix C, sections C.5.2 and C.5.3.

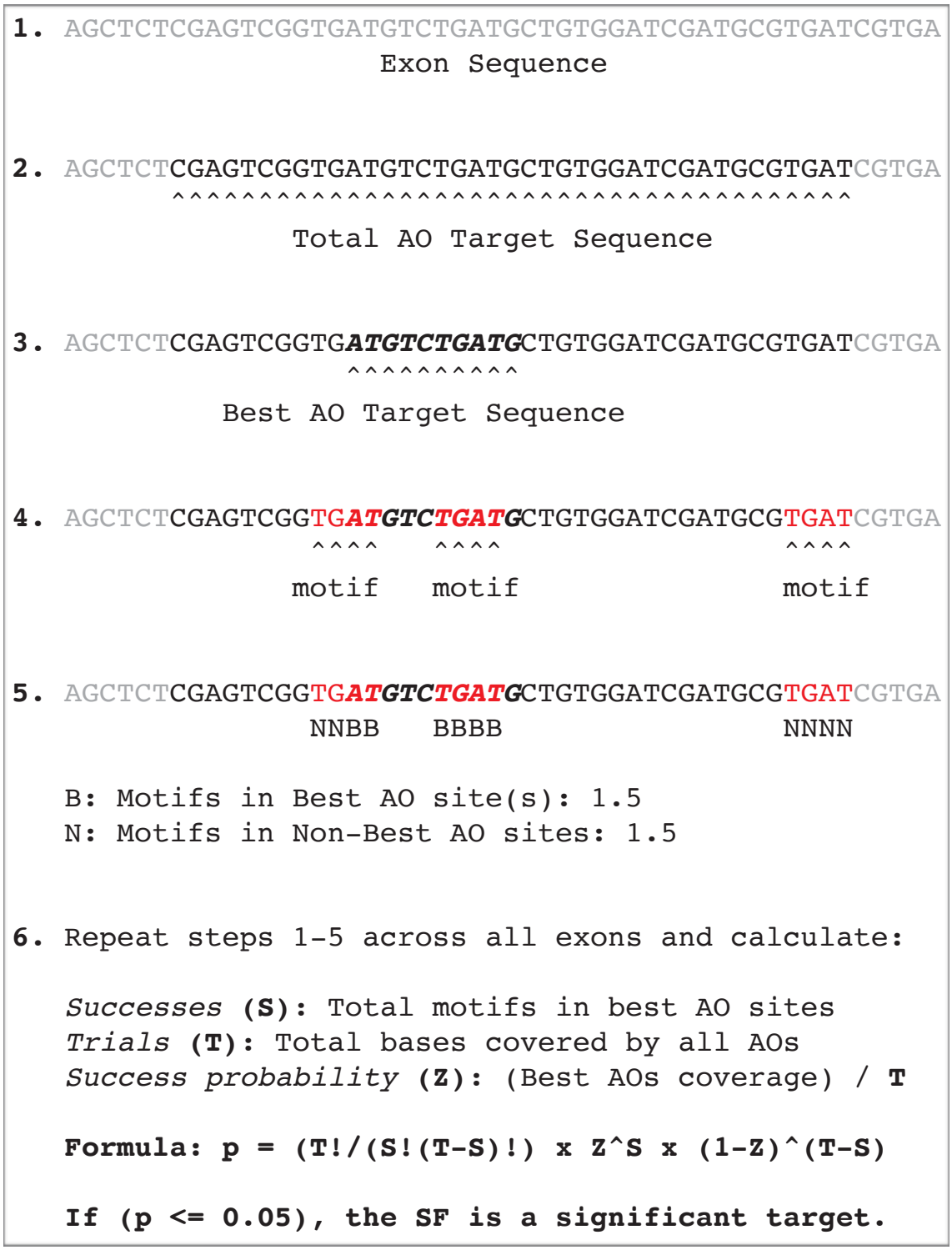


Figure 5.1: Concept diagram for calculation of splicing factor significance in AO efficacy. Starting with the sequence of a single exon (1), the total coverage of this exon by AOs is determined (2), and the target sequences of the best AO(s) identified within this region (3). Next, the motifs for the splice factor are identified (4) and tallied according to whether they fall inside or outside the target site of the best AOs (5). This process is repeated across all exons, and the totals and region sizes are used to calculate a binomial probability mass function that determines whether the splice factor is a significant target for the best AOs.

5.2.3 Statistics

The “Best” and “Rest” tallies for each splice factor were used as the basis for a two-tailed binomial test of probability density. The proportion of “Best” to “Rest” bases was used to calculate the expected number of splice factor sites in each group (e.g. if there were 1000 Rest bases and 200 Best bases, under the null hypothesis we would expect to see 200/1000 or 1/5 as many sites in the Best group as in the Rest group for any given splice factor). The test determined which (if any) AOs were significantly over- or under-represented in the most effective AOs.

5.3 Results

Splice Factor	# Sites in Best AO targets	# Sites in Non-Best AO targets	Total	P-Value	Strength	Target
MBNL1	20.83	37.00	57.83	0.0414	**1.35**	+
hnRNP H3	23.90	45.00	68.90	0.0435	**1.30**	+
SC35	30.68	57.86	88.54	0.0274	**1.30**	+
SRp40	57.89	111.11	169.00	0.0075	**1.29**	+
SRp20	26.38	51.90	78.28	0.0409	**1.26**	+
SRp30c	66.09	133.94	200.03	0.0084	**1.24**	+
hnRNP H2	49.87	108.17	158.03	0.0326	**1.18**	+
hnRNP H1	49.87	108.17	158.03	0.0326	**1.18**	+
ETR-3	12.80	50.67	63.47	0.0467	**0.76**	-
ZRANB2	11.83	50.00	61.83	0.0374	**0.72**	-
HTra2beta1	9.83	43.38	53.21	0.0359	**0.69**	-
SRp38	4.00	26.55	30.55	0.0178	**0.49**	-
RBM5	1.33	19.07	20.40	0.0148	**0.25**	-
CUG-BP1	0.40	11.39	11.79	0.0330	**0.13**	-

Table 5.1: Results of Splice Factor Significance Analysis. This table shows all 14 splice factors that were found to be significantly over- or under-represented in the best AO target sites. The “Strength” index shows the frequency of the splice factor in best AO sites compared to expected (eg. SRp30c sites coincide with best AO sites 1.24 times as often as would be expected to occur randomly). Splice factors were categorised as “+” or “-” targets depending on whether their strength index was greater than or less than 1.00, respectively.

5.4 Discussion

The original intent in detecting these splice factor biases was to facilitate building a predictive model for AO targeting. If achieved, such a model would enable more efficient oligomer design and reduce the timeline for the pre-clinical development of targeted genetic therapies such as exon skipping. Unfortunately, while some attempts at building such a model have performed better than random, none were noticeably better than the empirical approach currently used to target AOs - i.e. focusing on acceptor splice sites and SRp motif hotspots. It is likely that more information is need to create a model that can surpass this approach. The secondary structure of RNA is an important determinant of splicing (Yang *et al.* 2011) that was not accounted for in this study, and there may be other splice factors - or further binding motifs for existing splice factors - that have not yet been discovered.

Out of 71 splicing factors investigated, eight were found to be significantly targeted by the most effective AOs across all *DMD* exons, and six more were found to be significantly avoided. These results form the basis for a predictive model for AO-induced exon skipping, which is to our knowledge the first of its kind.

The goal of exon skipping AO therapy is to disrupt exon definition, such that the exon is not recognised by the spliceosome and is excluded from the mature transcript, along with with flanking introns. Originally, the most obvious targets were the splice sites themselves, and indeed AOs targeted to acceptor sites and (to a lesser degree) donor sites have a much higher efficacy than those targeted elsewhere; but they are not effective often enough for this to suffice as a design strategy. Members of the serine-rich protein (SRp) family are known to play important roles in exon definition, and the results of this study verify that four SR proteins in particular (SRp20, SRp30c, SRp40 and SC35) are significantly targeted by the most effective AOs. However, an SRp site on its own is only a weak indicator of a potential AO target, and there are no “exclusive” splice factors that occur always within or always outside of most effective AO target sites. This accords with what is already known about splicing

complexity and robustness, and it highlights why a predictive model must encompass as many factors as possible.

Serine and arginine rich proteins (SRps) are a ubiquitous protein family with roles in spliceosome assembly, and both constitutive and alternative splicing (Howard *et al.* 2015). There is evidence that these factors are especially important for exon definition (as opposed to intron definition) in large genes with high intron/exon ratios, a description that certainly applies to *DMD* (Berget 1995). Each SRp has its own unique characteristics and functions - for example, SC35 has established roles in mediating interactions between the U1 and U2 snRNPs, and in splice site definition (Fu and Maniatis, 1992). However, it is likely that their collective role here is simply that of exon definition. As the prime role of most exonic sequences is to encode functional polypeptide, this necessarily places severe limitations on the capacity of those same sequences to regularly present exon recognition motifs of a specific type and location. It is therefore not unusual that we see great variation in SRp variety, density and distribution from one *DMD* exon to another.

A notable exception to this trend was SRp38 which, rather than being targeted, was instead significantly avoided by the most effective AOs. However, prior research has indicated that SRp38 has an established role as a splicing repressor, making it somewhat the “black sheep” of the SRp family (Shin *et al.* 2005).

The hnRNP family is one of enormous variety: in the number of members, their functions, and the RNA motifs that they recognise. It is intriguing, then, that out of this entire family only the three members of the hnRNP H subgroup (H1, H2 and H3) would emerge as being significant as effective exon skipping AO target sites, and more so that they are positively targeted, since hnRNP Hs are generally characterised as splice silencers (Paul *et al.* 2006). A possible explanation is that the hnRNP Hs are being antagonistically silenced - that is, quite literally crowded out of their exonic binding sites - by SRps. This would create a positive correlation between the incidence of SRp and hnRNP H motifs and explain the relationship observed.

Perhaps the most interesting relationship revealed by this study is the positive target value of MBNL1. Eight splice factors in total were identified as positive targets, but while the other seven are all characterised as generalised splicing factors, MBNL1 is known to be a crucial component in the pathology of both forms of myotonic dystrophy (DM1 and DM2) (Jog *et al.* 2012). DM1 patients have a CTG trinucleotide expansion on one *DMPK* allele, and when it is transcribed, the CUG repeats in the mRNA form a sequestration site for MBNL1. This necessarily limits free MBNL1 available to act as an enhancer in the splicing of several other genes transcripts (Ranum *et al.* 2006; Jog *et al.* 2012), leading to errant exon skipping and a disease phenotype that is both complex and almost ubiquitous throughout the organs of the body. However, to date there has been little research into whether MBNL1 is involved in *DMD* splicing, though available data indicates that it does play a role (Klinck *et al.* 2014; Rau *et al.* 2015), at least in the splicing of exon 78, the second-last *DMD* exon. A 2007 study (Nakamura *et al.*) confirmed this and also found an increase in *DMD* e71 skipping in DM1 patients. While promising, these studies only examined a small fraction of the splice events that occur in the maturation of the DMD transcript. To our knowledge, this is the first *in silico* evidence of the role of MBNL1 in the splicing of the *DMD* transcript, and the first evidence that MBNL1 may be involved in the splicing of other *DMD* exons.

The CUG-BP1 (also called CELF1) and CUG-BP2 (ETR-3) binding motifs are both significantly avoided by effective exon skipping AOs, although much more so for CUG-BP1 than CUG-BP2. Intriguingly, like MBNL1 the CUG-BP proteins also play a key role in the splicing defects that lead to myotonic dystrophy (Timchenko *et al.* 2004; Lu *et al.* 1999), though the MBNL1 and the CUG-BP family appear to be antagonistic in their effects on splicing regulation (Wang *et al.* 2015; Kalsotra *et al.* 2008). This antagonism accords with MBNL1 being a strong positive AO target and CUG-BP1 being a strong negative target, as indicated in Table 5.1. It should be noted that the Kalsotra study examined the role of these factors in binding the flanking introns, whereas the most effective AOs in our dataset rarely targeted bases outside of the exons and splice sites. However, the presence of an intron-binding role for CUG-BPs in the splicing of exons in some genes does not preclude an exon-binding role in others. These results would appear to suggest that such a role exists.

RBM5 (RNA binding motif protein 5) has been shown to induce exon skipping in some transcripts, ostensibly by competing with the spliceosome for binding at the polypyrimidine tract (Jin *et al.* 2012). It therefore makes sense that RBM5 motifs are almost universally avoided by the most effective AOs. ZRANB2 is also implicated in the alternative splicing of numerous gene transcripts, acting to inhibit the inclusion in the transcript of the exons to which it binds (Yang *et al.* 2013).

The detection of hTra2-beta 1 motifs as a negative target is more puzzling. This protein has an established role as a promoter of inclusion of exon 7 in the *SMN2* gene transcript (Hoffman *et al.* 2000) and exons v4 and v5 in the *CD44* transcript (Watermann *et al.* 2006). Some hint of an explanation may be found in the latter of these two studies, which found SC35 acted antagonistically in relation to hTra2-beta 1 and diminished exon inclusion in *CD44,* despite SC35 more generally having an established role as an enhancer. This suggests that perhaps the roles of more than one splicing factor are 'inverted' in splicing of *DMD* transcripts.

It is difficult to draw conclusions about any splicing factors that are absent from Table 5.1. It is probable that many other factors do have either positive or negative value as AO skipping targets, but that their roles are too infrequent or too subtle to cross the threshold of significance in a study of this magnitude.

Exon-skipping antisense oligomer (AO) datasets present a unique challenge for statistical analysis. It is generally accepted that a reliable way of improving the definition and sensitivity of an experiment is to increase the sample size (practicality permitting), but this is only possible up to a point with exon skipping AO datasets. In our group's experience, for any single gene, there are rarely more than two unique, most effective AOs per exon; and since the number of exons is essentially fixed, this creates a hard upper limit on the sample size. Secondly, there is an enormous amount of work required to determine the most effective AO for even a single exon. Every AO that is designed requires multi-dose testing in human myoblast cultures, followed by RNA extraction, nested PCR and agarose gel verification - a process that must be scaled up or

repeated many times over to select the best exon skipping AOs for even a single exon. This means that, even when it is possible to expand the sample size, it is no simple matter to do so.

With these considerations in mind, two questions become crucial: (i) Can a predictive model built from the exon skipping AO dataset of one gene be usefully applied to a different gene? and (ii) Is it statistically valid to combine exon skipping AO datasets from multiple genes together into a "superset" to create a generalised predictive model?

It could be argued that the answer to the first question is "yes", but only in a limited sense. Eleven of the significant splicing factors identified in this experiment (all those in Table 5.1 except for MBNL1, CUG-BP1 and CUG-BP2) appear to be ubiquitously expressed. Therefore, it is reasonable to assume that they would function similarly for other human gene targets. A model such as this one, once pruned of splicing factors specific to the parent gene, might provide a good starting point for targeting AOs to a different, previously unexplored gene. Alternatively, if the gene is a closely related one, the existing model could be used as-is.

The second question is more difficult to answer at this time. It would certainly be worthwhile to at least attempt a generalised model, once enough data on other genes has been accumulated, though great care would need to be taken in its construction to ensure that it is not biased by genes with high exon counts, or by genes with significantly similar expression profiles. However, even this rests on the assumption that the same splicing factors function in the same way, regardless of gene identity or tissue specificity, and there is tentative evidence presented here (see section *5.4.8*) that this may not always be the case. A more prudent approach might be to restrict each generalised model to a single gene family wherever possible, as this would strike a balance between sample size and model specificity.

One of the original goals for this study was to reveal exon definition roles for splicing factors, both general and specific to myogenic cells. What was unexpected was that the identities of the involved specific splicing factors would

indicate a potential role for *DMD* transcript mis-splicing in myotonic dystrophy, perhaps to a much greater extent than previously thought. This, however, has yet to be empirically verified. It is therefore worthwhile investigating whether this relationship accords with what is already known about myotonic dystrophy, and how future investigations in this area might best proceed.

Schara and Schoser (2006) describe the symptoms of the myotonic dystrophies thusly:

> "Myotonic dystrophies (DMs) are autosomal dominant, multisystemic diseases with a core pattern of clinical presentation including myotonia, muscular dystrophy, cardiac conduction defects, posterior iridescent cataracts, and endocrine disorders."

Other symptoms can include cognitive impairment (Antonini *et al.* 2006), gastrointestinal dysfunction (Tieleman *et al.* 2008) and sleep disorders, such as hypersomnia and apnea (Shepard *et al.* 2012). Overall symptom severity and age of onset varies from patient to patient.

The two forms of myotonic dystrophy (DM1 and DM2) have similar genetic origins, albeit on different genes. DM1, the more common form of the disease, originates from a CTG trinucleotide repeat expansion in the *DMPK* gene (Aslanidis *et al.* 1992), whereas DM2 arises from a CCTG tetranucleotide repeat expansion in the *ZNF9* gene (Ranum *et al.* 1998). The pathologies of both forms of DM are very similar, and in both cases the range and severity of symptoms directly correlates to the magnitude of the expansion, while age of symptom onset negatively correlates to expansion size up to a point (Ho *et al.* 2015). Both expansion types are in non-coding regions and do not directly impede the function of the genes in which they reside, but as previously discussed, the expansions within the primary transcript act as sequestration sites for MBNL1 and thereby deplete the quantity of this factor that is available for modulating the splicing of other gene transcripts.

If, in addition to its other splicing roles, MBNL1 is also important to the splicing of multiple *DMD* exons, we would expect its deficiency to lead to an increase in

mis-spliced transcripts, with either exons missing or segments of intron included. However, given that even the most severe DM1 and DM2 phenotypes still have substantially better muscular function age-for-age than Duchenne patients, this leaves three possible explanations for how MBNL1 sequestration might affect *DMD* splicing: (a) correct *DMD* splicing can still proceed in the absence of MBNL1, but at a lower rate; (b) *DMD* splicing is chaotically disrupted by the absence of MBNL1, but some of the transcripts produced encode in-frame but truncated BMD dystrophin that retains some or most of the function of the full length product; or (c) some combination of these two effects.

A comparison of the symptoms of Becker muscular dystrophy and myotonic dystrophy shows substantial overlap, with patients affected by either disease exhibiting progressive global muscle weakness, cardiac involvement (Khalighi *et al.* 2015) and hyperCKemia (Finsterer *et al.* 2015). The range of symptoms for DM obviously eclipses that of BMD, but it may be that DM patients are experiencing what is effectively a mis-splicing induced form of BMD, that is at once masked and compounded by MBNL1-related splicing defects in other gene transcripts.

If this proves to be a valid explanation of DM then there are important implications for how it and BMD may be studied in future. DM and BMD would no longer be similar merely symptomatically but also molecularly, which would allow closer comparisons between the progress and patient outcomes of the two diseases.

The next step should be to verify experimentally if DM mutations do affect *DMD* pre-mRNA splicing. This could be done by culturing myoblasts from a patient with severe DM1, extracting RNA and performing RNAseq on the *DMD* transcripts. (Alternatively, AOs could be used to deplete MBNL1 in normal myoblasts, thereby simulating the pathology of DM). The results obtained could then be compared to the transcript diversity found in normal human myoblasts. If a relationship is evident, it could be investigated in more detail by similarly studying DM2 myoblast lines and those of patients with less severe variants of DM. Depending on the patterns found, it might be that some DM patients would benefit from exon skipping therapies originally developed for DMD sufferers.

AOs designed to enhance exon inclusion might prove a better option, if particular exons are pervasively excluded from DM-*DMD* transcripts, as these AOs would restore a dystrophin product that is full-length or near full-length. However, this type of AO is not nearly as thoroughly tested in *DMD* as exon-skipping AOs and so it is not nearly as close to being ready for *in vivo* evaluation.

A second possibility is to use AOs targeted to MBNL1 to deplete its translation

Preliminary work has indicated elevated native exon skipping in the central and 3' regions of the *DMD* transcript, in DM1 fibroblasts that have been transformed to myoblasts with ADV MyoD (data not shown). While the use of RNA from DM1 or DM2 patient muscle would be preferable, DM patients are typically diagnosed clinically, without undergoing muscle biopsy, and thus access to patient muscle samples has not yet been possible.

Whole transcriptome sequencing (WTS) or RNA-seq (Wang *et al.* 2009) could be a more comprehensive, if more expensive, approach to measuring the effect of DM genotypes on *DMD* transcript splicing. WTS works by converting purified total RNA extracts to cDNA and sequencing all the transcripts using NGS. Even if our hypotheses about *DMD* splicing in DM patients prove to be incorrect, this disease still presents as an excellent candidate for WTS generally. The core pathology of the disease is one of widespread splicing disruption, with the possible involvement of many as yet unidentified genes. A WTS comparative analysis of myotonic and normal muscle would provide an enormous breadth of unbiased data on the transcriptive phenotype of these diseases, with great potential for unexpected insights that could fuel future research. Of course, given how widespread the symptoms of DM are throughout the body, it would be necessary to perform WTS analyses on many other affected tissues before a complete map of the disease pathology could be obtained.

Chapter 6

General Discussion

6.1 Large intron deletion breakpoints can be efficiently mapped using Fractal PCR

As was discussed by Weise *et al.* (2012), deletions and duplications are often the asymmetrical products of the same type of event, namely non-homologous recombination. The work presented in Chapter 3 demonstrated how Fractal PCR could be a useful technique for localising deletion breakpoints in large introns, and the efficiency of Fractal PCR was compared to that of similar diagnostic strategies. It is worthwhile, however, to put aside these comparisons and consider the implications of what has been, and can be, revealed by the sequencing of whole exon mutation breakpoint junctions.

There is some consensus that one of the main genomic factors facilitating whole exon mutations is microhomology between the breakpoint junctions in the reference sequence (Hastings *et al.* 2009; Vissers *et al.* 2009; Verdin *et al.* 2013). However, as these microhomologies are often no larger than 2bp, this in itself offers little clue as to when or how the whole exon mutations may arise. The mechanisms behind these genetic lesions are thought to be a mix of otherwise beneficial repair mechanisms, such as non-homologous end-joining, fork-stalling-and-template-switching (Vissers *et al.* 2009), and microhomology-mediated break-induced replication (Hastings *et al.* 2009). At present, there appears to be no way of conclusively determining which mechanism brought about which mutation; and the origins of exonic deletion breakpoint insertions, such as those observed in Chapter 3 (see Figs. 3.14, 3.18 and 3.19), remain even more enigmatic.

Even if further research can reveal the precise origins of whole exon mutations, the question remains: Does the patient ultimately benefit? Currently there is no indication that such research could lead to any direct benefit for those with such genetic lesions, and while future treatment options may be tailored to the genome of the individual, in most cases they are more likely to reflect the mutation itself rather than how it came about. It is possible that altered intron composition could alter the efficacy of exon skipping AOs, but this has yet to be experimentally verified.

A possible clinical application for understanding whole exon mutation mechanics could be as a contribution to genetic counselling. In cases where a patient has a *de novo* whole exon mutation that does not exist in the somatic line of either parent, sequencing of both the patient's affected chromosome and the corresponding chromosomes in their parents should make it possible to match the mutated chromosome to its unmutated precursor(s). If this can be done often enough to build up a sufficiently large dataset, an analysis could be performed on the precursor mutation sites to search for common factors amongst them - as has already been done for potentially pathological gene variants linked to breast cancer (Foley *et al.* 2015). Knowledge of such factors could then be used to offer better genetic counselling to would-be parents on the *de novo* whole exon mutation risk that would be borne by their offspring.

6.2 Pseudoexons and other mis-splicing events occur at low levels within the transcript population of normal humans

Nearly all genes in the human genome undergo some degree of alternative splicing (Pan *et al.* 2008). Investigating transcript diversity is a worthwhile task that can have implications for genetic disease pathology (Bettencourt *et al.* 2013). However, understanding the full variety of spliceoforms that arise from even a single gene is a formidable task, one that is especially complicated by the existence of rare transcripts, such as those detected in Chapter 4, that blur the line between "true", functional transcripts with low copy number and those that are simply mis-spliced and awaiting nonsense mediated decay.

Halvardson *et al.* (2013) successfully used exome RNA capture sequencing to identify a large diversity of new splice junctions and exons, many of which had escaped detection via other methods, due to their rarity. Sultan *et al.* (2008) had similar success deep sequencing the transcriptome of human embryonic kidney cells and B cells. It remains unknown whether any of the rare transcripts discovered by these studies are translated into proteins, or have other distinct functions. Future studies may investigate whether the sequences of such cryptic exons are evolutionarily conserved, both within populations and across species, compared to the surrounding intron, as conservation would imply the exons retain function of some kind. Another approach would be to perform predictive

modelling of the putative protein isoforms encoded by rare transcripts, as this has proven useful in the past for determining protein function (Lee *et al.* 2007). Research in this area may also lead to important insights into the subtlety and complexity of spliceosome behaviour, and perhaps further reveal (Hsu and Hertel 2009) that many so-called splicing "errors" are in fact anything but.

6.3 Bioinformatic analysis of exon-skipping DMD *antisense oligonucleotides points to a stronger role for dystrophin mis-splicing in Myotonic Dystrophy*

Our group has done extensive work on developing exon-skipping antisense oligonucleotides for the treatment of DMD (Wilton 2013), but AOs are also being investigated as a treatment for a number of other genetic diseases, including neufibromatosis types 1 and 2, ataxia telangiectasia, and Niemann-Pick disease type C (Siva *et al.* 2014). Some work has been done to apply bioinformatics to the strategic design and selection of antisense oligomer target sites (Camps-Valls *et al.* 2004; Bo *et al.* 2006; Craig and Liao 2006; Pramono *et al.* 2012), though of these four only the latter one explicitly looked at exon-skipping AOs.

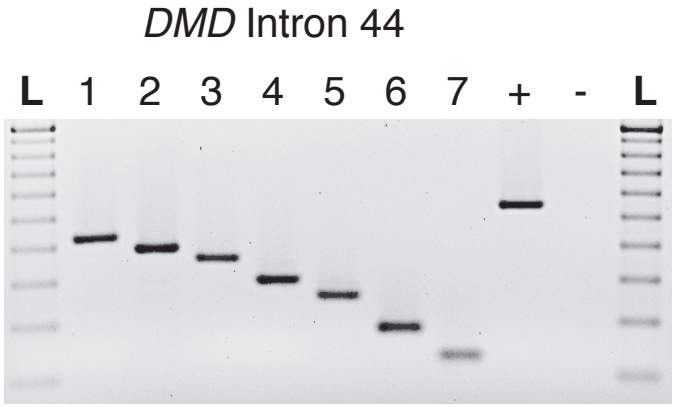
Efficient AO design is a pressing problem for all research groups investigating their use in personalised medicine, and there is little doubt that bioinformatic techniques will continue to yield improvements in this process. But regardless of how AOs have been designed, researchers should be aware that the records of their AOs' empirically demonstrated efficacies are an invaluable source of information, beyond their most obvious capacity to inform selection of future AO targets. This was clearly demonstrated by the work discussed in Chapter 5, and data-mining of other biological data has already yielded positive results towards the treatment of cysticercosis (Yan *et al.* 2014) and schistosomiasis (Bos *et al.* 2009), and uncovered a potentially stronger link between two genes implicated in sporadic amyotrophic lateral sclerosis (Kitano *et al. 2015)*. But however rigorous bioinformatic predictions may be, there always remains the need to empirically verify the results generated from such studies. The prediction made in Chapter 5 is no exception.

6.4 Closing Remark

As a disease gene, much of the research on *DMD* is rightly focused upon manipulating its expression in order to ameliorate the symptoms of the diseases caused by its dysfunction. However, investigation of the baseline nature of *DMD* and its transcripts, in both the gene's normal and myriad mutated forms, is equally important. Not only does such research provide a solid grounding for the development of new therapies, but it can also reveal unexpected patterns that have the potential to be transformative in how we view the dystrophinopathies and other genetic disorders.

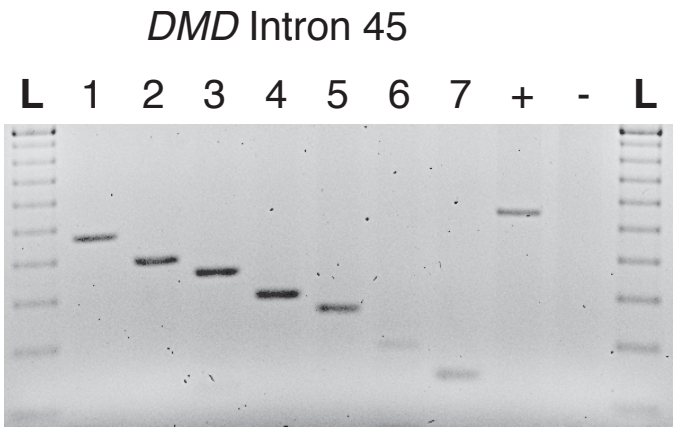
Appendix A

Additional Data for Chapter 3: Evaluations of intronic primer pairs



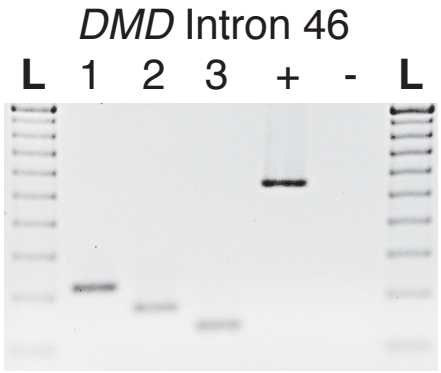
Lane	F Primer	R Primer	Size (bp)
1	5867	5868	**423**
2	5869	5870	**399**
3	5871	5872	**365**
4	5873	5874	**303**
5	5875	5876	**261**
6	5877	5878	**192**
7	5879	5880	**140**
+	5742	5743	**557**
-	None	None	**0**

Figure A.3.1: Phase I evaluation of primer pairs targeting *DMD* intron 44, using normal human gDNA as template. Amplicons are ordered according to the position of the their intronic target site, 5' to 3' left to right. Positive control uses previously verified primer pair targeting *DMD* intron 47, and negative control uses DNA but no primers. "L" indicates 100bp ladder size standard.



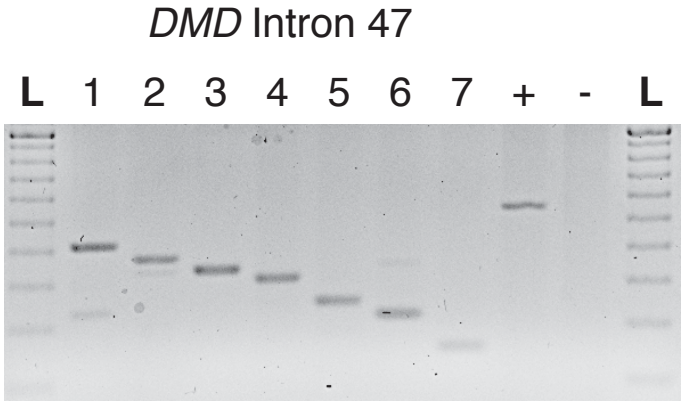
Lane	F Primer	R Primer	Size (bp)
1	6015	6016	**474**
2	6017	6018	**396**
3	6019	6020	**362**
4	6021	6022	**300**
5	6023	6024	**271**
6	6025	6026	**194**
7	6027	6028	**145**
+	5742	5743	**557**
-	None	None	**0**

Figure A.3.2: Phase I evaluation of primer pairs targeting *DMD* intron 45, using normal human gDNA as template. Lane 7's product was faint but detectable. Amplicons are ordered according to the position of the their intronic target site, 5' to 3' left to right. Positive control uses previously verified primer pair targeting *DMD* intron 47, and negative control uses DNA but no primers. "L" indicates 100bp ladder size standard.



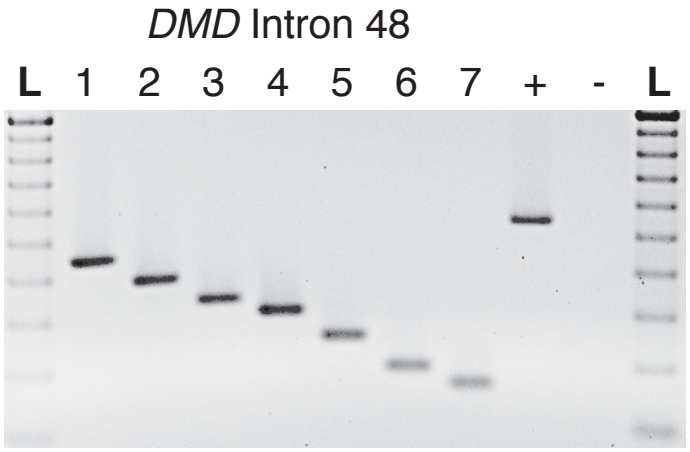
Lane	F Primer	R Primer	Size (bp)
1	5881	5882	**226**
2	5883	5884	**176**
3	5885	5886	**141**
+	5742	5743	**557**
-	None	None	**0**

Figure A.3.3: Phase I evaluation of primer pairs targeting *DMD* intron 46, using normal human gDNA as template. Only three primer pairs were required for intron 46 as, at 2334bp in length, it is drastically smaller than the other introns in this region. Amplicons are ordered according to the position of the their intronic target site, 5' to 3' left to right. Positive control uses previously verified primer pair targeting *DMD* intron 47, and negative control uses DNA but no primers. "L" indicates 100bp ladder size standard.



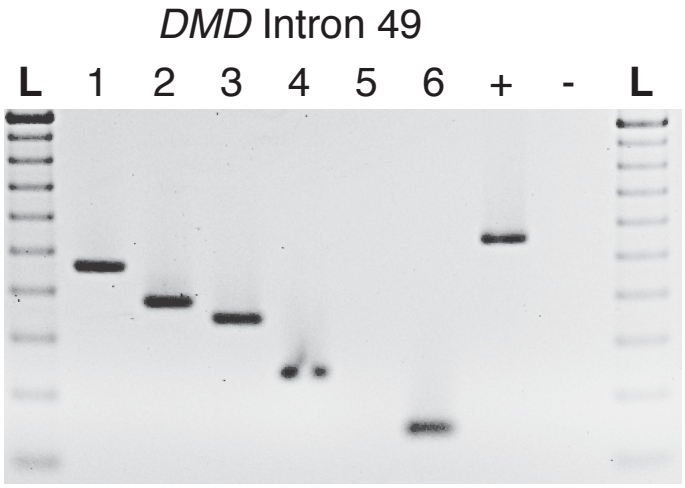
Lane	F Primer	R Primer	Size (bp)
1	6029	6030	**421**
2	6031	6032	**379**
3	6033	6034	**342**
4	6035	6036	**315**
5	6037	6038	**254**
6	6039	6040	**224**
7	6041	6042	**157**
+	5742	5743	**557**
-	None	None	**0**

Figure A.3.4: Phase I evaluation of primer pairs targeting *DMD* intron 47, using normal human gDNA as template. Non-specific products were visible in lanes 1, 2 and 6, possibly due to mispriming, however it was determined they were below the level necessary to be mistaken for on-target products in the multiplex. On-target amplicons of the expected sizes predominated in all seven lanes. Amplicons are ordered according to the position of the their intronic target site, 5' to 3' left to right. Positive control uses previously verified primer pair targeting *DMD* intron 47, and negative control uses DNA but no primers. "L" indicates 100bp ladder size standard.



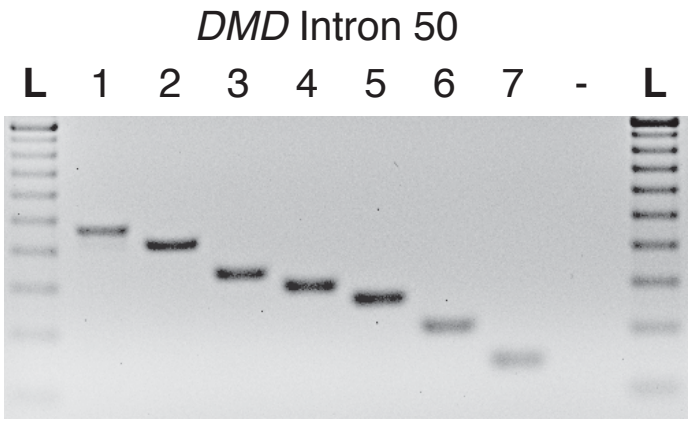
Lane	F Primer	R Primer	Size (bp)
1	5887	5888	**453**
2	5889	5890	**389**
3	5891	5892	**336**
4	5893	5894	**306**
5	5895	5896	**253**
6	5897	5898	**197**
7	5899	5900	**167**
+	5742	5743	**557**
-	None	None	**0**

Figure A.3.5: Phase I evaluation of primer pairs targeting *DMD* intron 48, using normal human gDNA as template. Amplicons are ordered according to the position of the their intronic target site, 5' to 3' left to right. Positive control uses previously verified primer pair targeting *DMD* intron 47, and negative control uses DNA but no primers. "L" indicates 100bp ladder size standard.



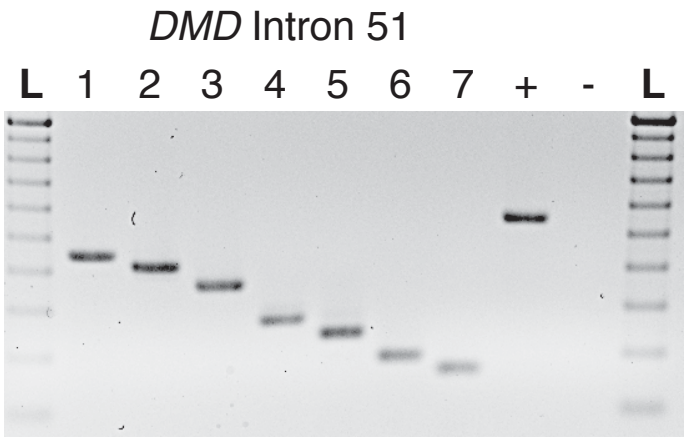
Lane	F Primer	R Primer	Size (bp)
1	5901	5902	**470**
2	5903	5904	**389**
3	5905	5906	**329**
4	5907	5908	**246**
5	5909	5910	**174**
6	5911	5912	**149**
+	5742	5743	**557**
-	None	None	**0**

Figure A.3.6: Phase I evaluation of primer pairs targeting *DMD* intron 49, using normal human gDNA as template. Intron 49 is a smaller intron (16649bp) and has large tracts of microsatellite repeats. For these reasons only six primer pairs were targeted to this intron. The reaction in lane 5 failed to produce a product in repeated PCRs, due to either inefficient design or the presence of a small deletion at one of the primer sites. It was not included in any following multiplex reactions. Amplicons are ordered according to the position of the their intronic target site, 5' to 3' left to right. Positive control uses previously verified primer pair targeting *DMD* intron 47, and negative control uses DNA but no primers. "L" indicates 100bp ladder size standard.



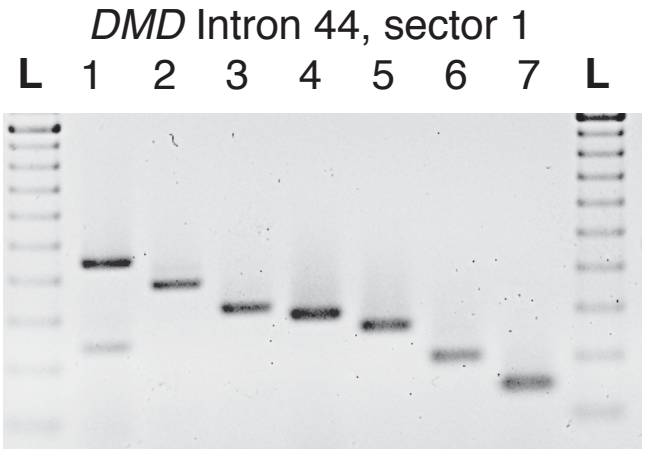
Lane	F Primer	R Primer	Size (bp)
1	5913	5914	**474**
2	5915	5916	**416**
3	5917	5918	**336**
4	5919	5920	**296**
5	5921	5922	**274**
6	5923	5924	**212**
7	5925	5926	**150**
-	None	None	**0**

Figure A.3.7: Phase I evaluation of primer pairs targeting *DMD* intron 50, using normal human gDNA as template. Amplicons are ordered according to the position of the their intronic target site, 5' to 3' left to right. Positive control uses previously verified primer pair targeting *DMD* intron 47, and negative control uses DNA but no primers. "L" indicates 100bp ladder size standard.



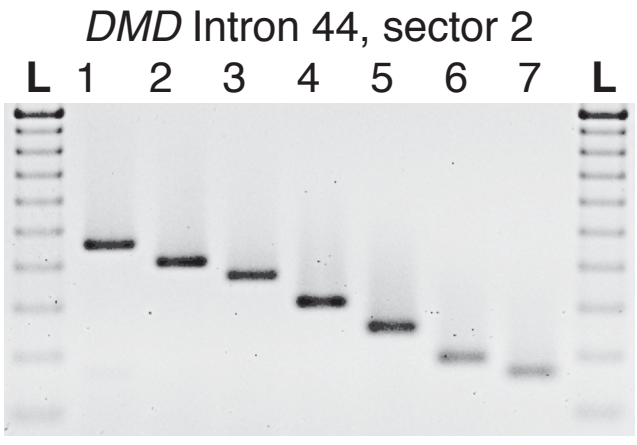
Lane	F Primer	R Primer	Size (bp)
1	5927	5928	**453**
2	5929	5930	**423**
3	5931	5932	**369**
4	5933	5934	**282**
5	5935	5936	**248**
6	5937	5938	**199**
7	5939	5940	**172**
+	5742	5743	**557**
-	None	None	**0**

Figure A.3.8: Phase I evaluation of primer pairs targeting *DMD* intron 51, using normal human gDNA as template. Amplicons are ordered according to the position of the their intronic target site, 5' to 3' left to right. Positive control uses previously verified primer pair targeting *DMD* intron 47, and negative control uses DNA but no primers. "L" indicates 100bp ladder size standard.



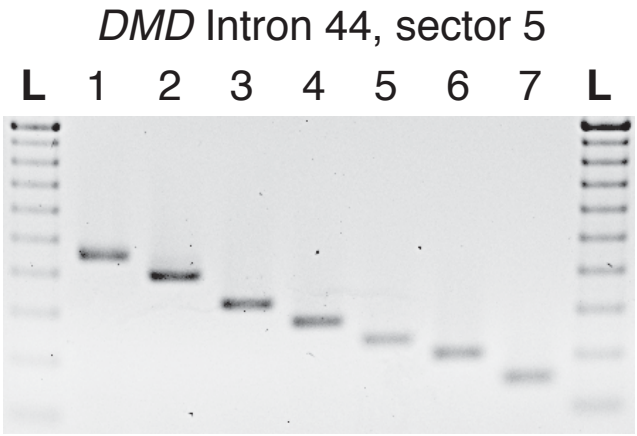
Lane	F Primer	R Primer	Size (bp)
1	6327	6328	**442**
2	6329	6330	**388**
3	6331	6332	**324**
4	6333	6334	**307**
5	6335	6336	**276**
6	6337	6338	**212**
7	6339	6340	**158**

Figure A.3.9: Phase II evaluation of primer pairs targeting *DMD* intron 44, sector 1, using normal human gDNA as template. The reaction in lane 1 produced an off-target amplicon of interfering size and was excluded from future multiplex PCRs. Amplicons are ordered according to the position of the their intronic target site, 5' to 3' left to right. Positive control uses previously verified primer pair targeting *DMD* intron 47, and negative control uses DNA but no primers. "L" indicates 100bp ladder size standard.



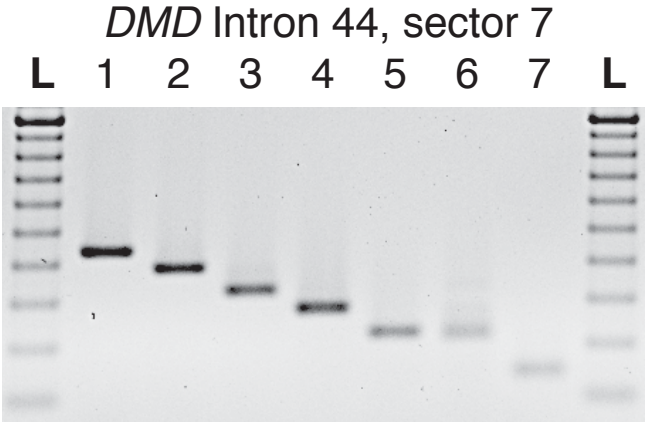
Lane	F Primer	R Primer	Size (bp)
1	6219	6220	**462**
2	6221	6222	**409**
3	6223	6224	**374**
4	6225	6226	**314**
5	6227	6228	**255**
6	6229	6230	**201**
7	6231	6232	**175**

Figure A.3.10: Phase II evaluation of primer pairs targeting *DMD* intron 44, sector 2, using normal human gDNA as template. Amplicons are ordered according to the position of the their intronic target site, 5' to 3' left to right. Positive control uses previously verified primer pair targeting *DMD* intron 47, and negative control uses DNA but no primers. "L" indicates 100bp ladder size standard.



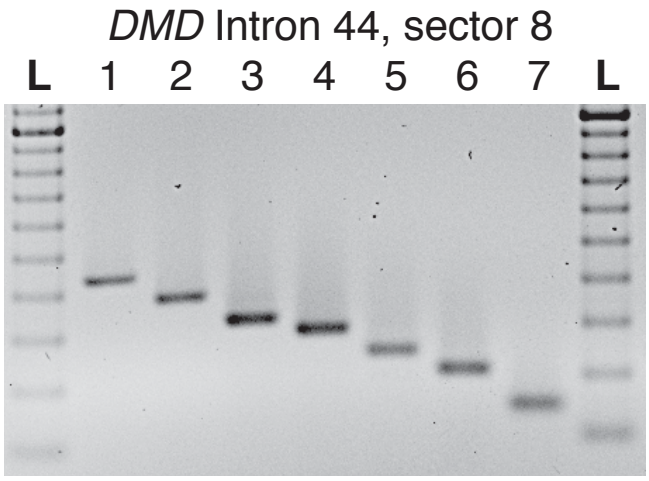
Lane	F Primer	R Primer	Size (bp)
1	6233	6234	**459**
2	6235	6236	**404**
3	6237	6238	**331**
4	6239	6240	**286**
5	6241	6242	**245**
6	6243	6244	**213**
7	6245	6246	**164**

Figure A.3.11: Phase II evaluation of primer pairs targeting *DMD* intron 44, sector 5, using normal human gDNA as template. Amplicons are ordered according to the position of the their intronic target site, 5' to 3' left to right. Positive control uses previously verified primer pair targeting *DMD* intron 47, and negative control uses DNA but no primers. "L" indicates 100bp ladder size standard.



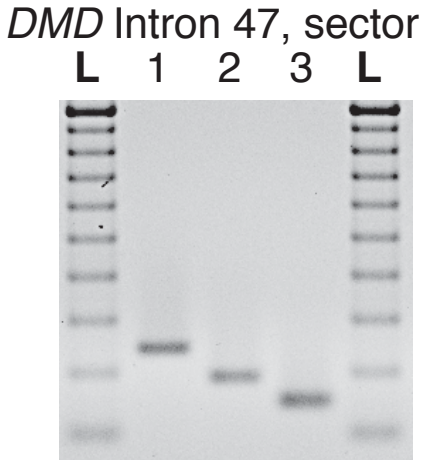
Lane	F Primer	R Primer	Size (bp)
1	6247	6248	**450**
2	6249	6250	**398**
3	6251	6252	**335**
4	6253	6254	**288**
5	6255	6256	**234**
6	6257	6258	**222**
7	6259	6260	**150**

Figure A.3.12: Phase II evaluation of primer pairs targeting *DMD* intron 44, sector 7, using normal human gDNA as template. Although the amplicons in lanes 5 and 6 are close in size (234bp and 222bp respectively), they are still distinctive and are included in the multiplex reactions. Amplicons are ordered according to the position of the their intronic target site, 5' to 3' left to right. "L" indicates 100bp ladder size standard.



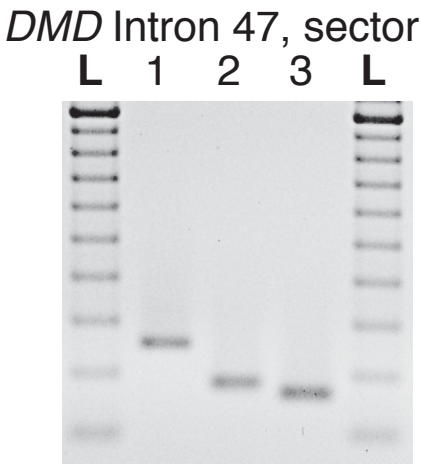
Lane	F Primer	R Primer	Size (bp)
1	6261	6262	**441**
2	6263	6264	**376**
3	6265	6266	**325**
4	6267	6268	**298**
5	6269	6270	**251**
6	6271	6272	**218**
7	6273	6274	**150**

Figure A.3.13: Phase II evaluation of primer pairs targeting *DMD* intron 44, sector 8, using normal human gDNA as template. Amplicons are ordered according to the position of the their intronic target site, 5' to 3' left to right. "L" indicates 100bp ladder size standard.



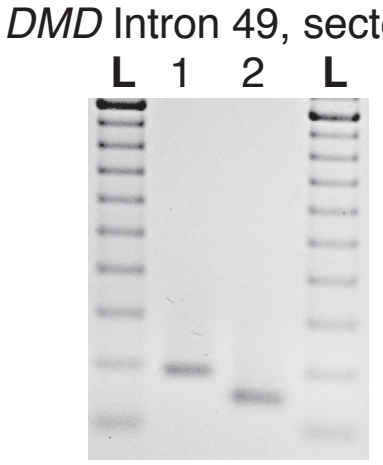
Lane	F Primer	R Primer	Size (bp)
1	6275	6276	**249**
2	6277	6278	**199**
3	6279	6280	**158**

Figure A.3.14: Phase II evaluation of primer pairs targeting *DMD* intron 47, sector 2, using normal human gDNA as template. Amplicons are ordered according to the position of the their intronic target site, 5' to 3' left to right. "L" indicates 100bp ladder size standard.



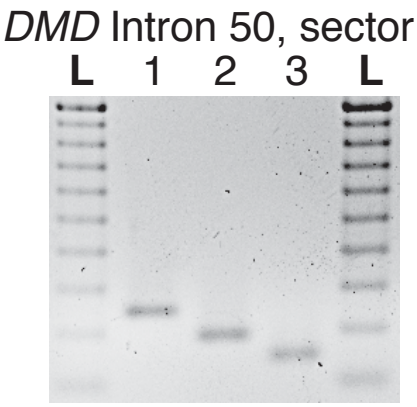
Lane	F Primer	R Primer	Size (bp)
1	6281	6282	**265**
2	6283	6284	**192**
3	6285	6286	**174**

Figure A.3.15: Phase II evaluation of primer pairs targeting *DMD* intron 47, sector 6, using normal human gDNA as template. Amplicons are ordered according to the position of the their intronic target site, 5' to 3' left to right. "L" indicates 100bp ladder size standard.



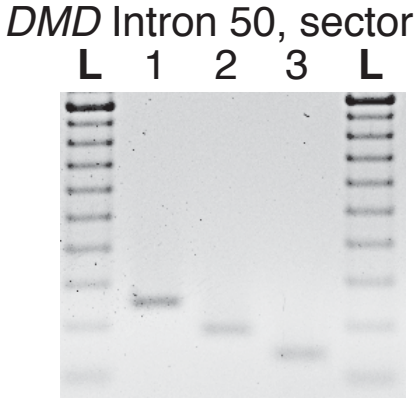
Lane	F Primer	R Primer	Size (bp)
1	6287	6288	**198**
2	6289	6290	**155**

Figure A.3.16: Phase II evaluation of primer pairs targeting *DMD* intron 49, sector 5, using normal human gDNA as template. Amplicons are ordered according to the position of the their intronic target site, 5' to 3' left to right. "L" indicates 100bp ladder size standard.



Lane	F Primer	R Primer	Size (bp)
1	6291	6292	**252**
2	6293	6294	**298**
3	6295	6296	**158**

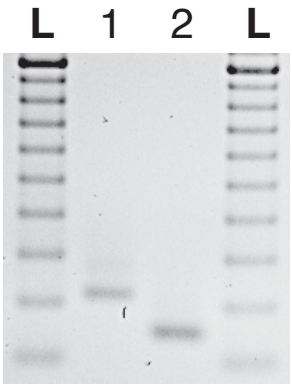
Figure A.3.17: Phase II evaluation of primer pairs targeting *DMD* intron 50, sector 5, using normal human gDNA as template. Amplicons are ordered according to the position of the their intronic target site, 5' to 3' left to right. "L" indicates 100bp ladder size standard.



Lane	F Primer	R Primer	Size (bp)
1	6297	6298	**255**
2	6299	6300	**195**
3	6301	6302	**151**

Figure A.3.18: Phase II evaluation of primer pairs targeting *DMD* intron 50, sector 7, using normal human gDNA as template. Amplicons are ordered according to the position of the their intronic target site, 5' to 3' left to right. "L" indicates 100bp ladder size standard.

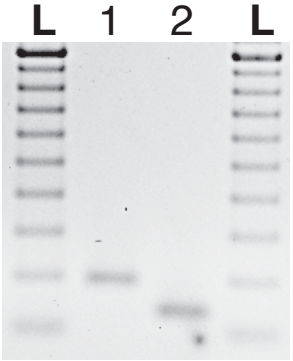
DMD Intron 51, sector 3



Lane	F Primer	R Primer	Size (bp)
1	6303	6304	**225**
2	6305	6306	**156**

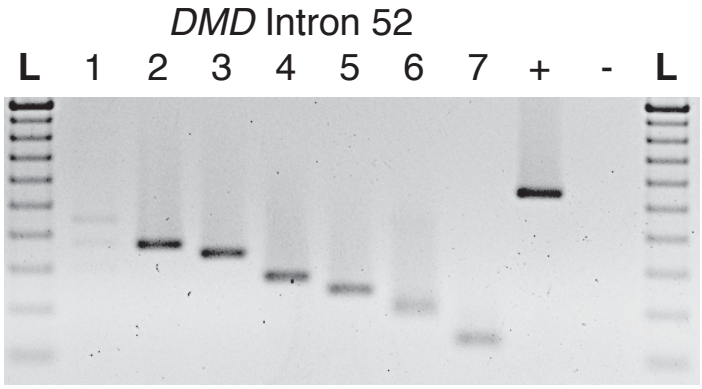
Figure A.3.19: Phase II evaluation of primer pairs targeting *DMD* intron 51, sector 3, using normal human gDNA as template. Amplicons are ordered according to the position of the their intronic target site, 5' to 3' left to right. "L" indicates 100bp ladder size standard.

DMD Intron 51, sector 7



Lane	F Primer	R Primer	Size (bp)
1	6307	6308	**201**
2	6309	6310	**143**

Figure A.3.20: Phase II evaluation of primer pairs targeting *DMD* intron 51, sector 7, using normal human gDNA as template. Amplicons are ordered according to the position of the their intronic target site, 5' to 3' left to right. "L" indicates 100bp ladder size standard.



Lane	F Primer	R Primer	Size (bp)
1	6313	6314	**457**
2	6315	6316	**375**
3	6317	6318	**340**
4	6319	6320	**282**
5	6321	6322	**248**
6	6323	6324	**215**
7	6325	6326	**143**
+	5742	5743	**557**
-	None	None	**0**

Figure A.3.21: Phase I evaluation of primer pairs targeting *DMD* intron 52, using normal human gDNA as template. This primer set was designed after the other Phase 1 sets, though ultimately it was not needed for any experimental work. The reaction in lane 1 failed to produce a distinct product and was excluded from future cocktails. Amplicons are ordered according to the position of the their intronic target site, 5' to 3' left to right. "L" indicates 100bp ladder size standard.

Appendix B

Additional Data for Chapter 4: Genomic Sequences of Targeted Pseudoexons

ID	Pseudoexon Sequence	Reference
DMDpe1a	aaaacagatatgaagaaatccacagcaacatttattataaaattactctttctcttccttggttttgcgg**CTTCTCGAGTTCATAGGAGACTTTCAGTTTCCAGTGACCTGGAAACTCACCATTCCTCATCACCATCCTTTACTGTAGTAACTTCCTTTTACCTGACCACCCTGCATAGTCACAGAAGATGCACTCCTGACAAGTGATCCTCAAAACAG**gtagtaatctctttgagaagtgtgaaatgtcttttgttctccttgtcgttcccatgattgtcagctgtga	Beroud *et al.* 2004
DMDpe11a	catcaatgctttgactggacagaaaaatacatctaacaaatatggctcatctctttttttttcccccaag**TGTCTATTTGACTCTGGAAATAAGATGGCATATGTGAGAGTGGATAGAGAAAAGGAGGTGCTGAACAAATGAGTGGGTTATTTTCTCCCAAGTGCAGTGAAGTTTGCGTATGTGAATATATGAATGAATAGATGAACAACTAAATTAAGTTATTCAG**gtaattccaatcttggaattgaattcgttgcagtaaatactttggaggtgcacaatctcaaactcctaat	Malueka *et al.* 2012
DMDpe18a	aatatttctcaaattgtgggtatttgtgctcatagtatctttttttttttttttttccacctgccttag**TGGAAGAGGCTATCCTTCACCTGGTTGAGGCTCATCTCTGGGTGTGTGTTCTCAGCAGCATCACTGACTATGTATTAAGCCACCTGGTTCCATTCAGCTGTATATCCAGATTGTCAAAAATCTACATCCCAG**gtctttatcattagcttttaaaccggttgtgttttcatcttttagaacgtgtcctctcctataaacatgc	Zhang *et al.* 2007
DMDpe25a	gatccaagctatgttctgatctccattttctaatcttagaatattatatgtatctgtgctttttcctacaa**GTATCACTCTGGCCATGTTCTGACTTTGTAGCCAAATGAGTTAGGTTGTAAAAGGAAGGAACAATGGCGCTCAAGGAGAAGAAGAAGACGATGCG**gtaaaaacaaggaagccatatgtgaatattgttaccaattcagcattccagagagaataatggaaatgaa	Tuffery-Giraud *et al.* 2003
DMDpe25b	atccaagctatgttctgatctccattttctaatcttagaatattatatgtatctgtgctttcctacaag**TATCACTCTGGCCATGTTCTGACTTTGTAGCCAAATGAGTTAGGTTGTAAAAGGAAGGAACAATGGCGCTCAAGGAGAAGAAGAAGACGATGCGGTAAAAACAAGGAAGCCATATGTGAATATTGTTACCAATTCAGCATTCCAGAGAGAATAATGGAAATGAAGTGTAAATCTATGCATTACAGAAATATCTACAGACAAA**ataagtgtgtgatacactcttttgggttaaagataattccttttttgtcaccagatgccaatgtctgcga	Ikezawa *et al.* 1998
DMDpe26a	ttatcagaatgtgagagctagaaggtactttgaaacttgtgtaattcacacttctctctctgcttatcag**AAAACTGCAATTCCCAGATAGGTCAAATGATTTAGCCATAGTCACAGACTTTATTTGTGGTAGAGCCCACAGGATTGAAG**gtattttattctatttcatctctttttttccttttcctttcctttccttttccttttccttttccttttcc	Trabelsi *et al.* 2014
DMDpe45a	actatggtttcagtcattttaatgtgcagtcactatcttcatacactccttttcttctggagtattctag**GAGAAGACATACCAGTCGAGGGGTTCTGGGGAGCCAGGCCTTCAAGCAATGGATTGCTGACAACATAATGAAGAGGATTTTACTTAGAATAATGTCAGTTGATAAAAGTTTGAATGGGAGACGGAAGCAAGGCAGTG**ggaagtggaattcctaaattgaggaacctctgaatcataatccttagcaataataattaagatttcaaaa	Gurvich *et al.* 2007
DMDpe56a	cagtactagcatgagtcacatgaaaacgaagtgtttttcattagtcactgcatccaatttttctaccag**AAATATTAAGAATTGTTGACTACAACAGTATGGAAAAGCAATAGATTCCAGTGTGTATTTCATGCCAAAAGTCTCAGCATTCTGCATGTGGAAATAAACATATGGCTAAACACTGCCTTTTCTCAAAATTGCCATCAAACTATCCTCTGTTTTGTGGCTCTCAAAA**gtaagtagccagatttttattcaagactgctttgtagttcacaataggtttattgtacttttcaatgga	Khelifi *et al.* 2011
DMDpe60a	aattgggtcacatgcccacctaggctaggggaatggaatcgtggaattgattgattgattattgttgcag**GTTGAGTCCTCCAAGAAGCAGATGCCAGGGCAGATTTATTGCAGAATAACACCTGTGAAGAAATAGGGGTGGAAGCAGAATTGAACAAG**ggaagccatcagatcacaatgcagatcttctagactctgtgtcagcacaacaaggaatgccagagcaaag	Beroud *et al.* 2004
DMDpe65a	acattagttttgataagagtttgtattcaggtttcccctcaaatatattttggcactgttttctttgcag**ATGACATGTGAATGCATTCTGAATGTATAACTTCCTTCTACCTGACTGAAAAGTATTTGGTGACAATTTTAACTCCTTGAAGACCTGAGTTGCTGTATAAAGTGGATTTGTTAAATTTTGATCTACCTTTTCTTAAGAGGGAGAAAG**ataagaaaattttccagtgataggaatttgttttagtccattagtgcccaggatttctttttgactgaggg	Deburgrave *et al.* 2007
DMDpe77a	tggaggtagtagggatggaaagaagaaatgacaaattccttgccttcaaggtacttatggtcaattgcag**ATACAGAACTCCAAAGAAATTCAAATCACAGTACAATCATTGATTACTGTATGATCCTGGCCCCTTGAACTGGCAGCTTGCTTACCTGCTCTGGAAAGTTGCTGGCTGCCCTGTTGGCACCCTGGGCATTTCTTCCACATCTAAACAAGAG**gttagtagagaagaagctacagcaaggctcatgactctgcttcttcataggcagccttgacttcctagac	Zhang *et al.* 2007

Table B.4.1: Genomic sequences of targeted pseudoexons in the human *DMD* gene.

Sequences were extrapolated from coordinates and sequence samples given in the referenced works, based on the current reference sequence for the *DMD* gene at time of writing (NG_012232.1).

ID	Pseudoexon Sequence	Reference
NF1pe3a	ttttcattaaatttactatttaaaaagtaatgctatttatattattgaacaattgttttattccttcag**TGATTTTACAGATGTTTCTTGATTCTTACTTTTAAATCAAAGTAAATTGAGGCAGTTTAAGAAGAACGGAGCTCACAAGTATACGTTTGTTATACATAGTTGGAATAT**ttaagtgaaagaaaaatacttaccaatatgttatgagtgcaattttgtagaatttgaggaatcttaagta	Pros *et al.* 2009
NF1pe6a	gggttttactgaaatttcaatatgcagtttgttttgcaaatatttttttttttcatttttgctgcttttg**GAGAAATACTCTTAGCTGGCTGAACATGAGATGGTTTCCTTCATGGTTTGCCTGTAGATGGAGATGTTCAGAATCAGCTGAATGATTTATTACTTTGTAGAAGTATTGGACCGGGAATTTATACATTAAAGG**gtaaataaatgaaccaggtaatccttgagagtctgtgattgcgtgttactccacttaggtgtatatctca	Pros *et al.* 2008
NF1pe6b	tcacaaaacatactttgtatgatttcagtttctttaaatgtattaagacttattttgtagcctaacatag**GTCTGTTTTGGAGAATGTTCCATGTACACTTGAGAAGAATATATTGTGCTGGTGTTGG**gtgagtttagtttttgacatgttacatttgagcattcatttaggagaaaagccaaaggaagacatttgtt	Messiaen and Wimmer 2008
NF1pe10a1	tctataattgaaacatattttctgtaagtttacctttctgagttctttctgttttacaaatgttcttcag**GCAAAATGCATGCATTCTGTAACTTAAAAGAGGTGGAGTGCCGTCGAAGGAAAAAGCAAATTTTACAATTAAACTA**gtaaatacaaagattgcctacatgtttcattaattctaagaagcacttatttttactcttctctaaaatt	Pros *et al.* 2008
NF1pe10b1	tattaataagatacaatttaacactaattaatgcaccaaatgtaatagttctatattgtgtctttcatag**GATGACATGTTTAACCTTTGTTGAGCTTCTTCAGTCCCTGGAGAGCAGCATCAA**gcaaggtttcttatcgttttgctcagtaactggctttttaaaaagtgttaaggtgctttcgattttggtg	Pros *et al.* 2008
NF1pe19a1	aaaaaatctgtacatgttgagtacagatgaaaccctccttttttttccccaaaatattttcaatccatgg**GTGGTTGAATCCATGGATATAGAACTCACTGATACAGAGGGCCCATTGTACATGCTTCTGATTTAAGGTAGCCATTTTGCCAAGATTACTTTGTAGAAA**gtaagtattaccttctccccatttgagatgattttgtattctgggatctgcatattaactcaaattattt	Fernandez-Rodriguez *et al.* 2011
NF1pe23a1	ccaaaggcaatttgtgggcatttgttgcattggatatccttggtctaattttattattgttacttttaa**ATTATAAATGAATGCAAAGAAACTTAATTTCAAGGCCTTAGGAAACGCTGACTGTCTTCATTCTTGCTTCTTGTTTGAAG**gtaatgtgagtggtttcttttcccagagatgacagtgtttcttaattgttagaagtatatggtgggaaag	Kannu *et al.* 2013
NF1pe30a	aagactttcacttacatttttacttttttttcctcttctgattttatatctgtggtatcctgtaactgaag**GAACTCTGAAGGAACTTTTTGGTAGGCCACATTGAGAAATCCACCAAAATTCTGATAATAAAAGCAAAGCGGCAAGAGTTTGGTACTTTACACACCTCTGGACCACTTACAAAGTATTTTCTTCACTGTTGCAGTTTAATTAGCTTCAGCATTTCTCTTCATAGCAGAAAAGTCCAC**ataagtatccatgttgcctccttaattttagagaactagtcattttctcttttcttaggttgactttga	Perrin *et al.* 1996
NF1pe30b	acagttttcattttgtggtgatgctttccttttaccaaactttctatgattaccacatttcctttataa**TGAGAATAAAACAACTTTTTAACAAGAAAGGACTAAAATGGAGGAAAATAAGACAAAACTTTTCAAAAATTGGCTTACTGGCTTTTAAAATTACTTTCTTCAAGGACTGTTCTTTCTTCGCCTCTACAAAAATATATTTGCCAAGTGTCTTTTCTCCAGGCCTGATTCTAG**gtaatagtctttacctttaccatttttccccgaattctttatgttaaataattgttgatgtgattttc	Perrin *et al.* 1996
NF1pe45a	tgatctgactcagcattttaaaagcactgattaacattaaatttaatttttttctttttgttttgcacag**ATCCTAAATTTAGTTTTAACTTCCTAAGCGCATGTCAGTATACAACAGATGGAAATAGTACTAAAACATG**ctaagtagcagacagagccaaccttgtcttaagcaaacatttaccgtatatggttacacatggttatatg	Pros *et al.* 2009

Table B.4.2: Genomic sequences of targeted pseudoexons in the human *NF1* gene.

Sequences were extrapolated from coordinates and sequence samples given in the referenced works, based on the current reference sequence for the *NF1* gene at time of writing (NM_001042492.2).

Appendix C

Additional Data for Chapter 5

Exon	Size (bp)	Total AOs	Best AOs	Exon	Size (bp)	Total AOs	Best AOs	Exon	Size (bp)	Total AOs	Best AOs
1	239	N/A	N/A	**27**	183	3	1	**54**	155	5	1
2	62	7	1	**28**	135	3	1	**55**	190	11	2
3	93	17	1	**29**	150	3	1	**56**	173	12	1
4	78	7	1	**30**	162	3	1	**57**	157	10	2
5	93	12	1	**31**	111	9	1	**58**	121	7	2
6	173	10	1	**32**	174	6	1	**59**	269	15	1
7	119	10	1	**33**	156	4	2	**60**	147	14	2
8	182	14	1	**34**	171	15	1	**61**	79	7	1
9	129	3	1	**35**	180	5	1	**62**	61	11	1
10	189	12	1	**36**	129	11	1	**63**	62	6	1
11	182	13	1	**37**	171	3	1	**64**	75	5	1
12	151	8	1	**38**	123	6	1	**65**	202	9	3
13	120	4	1	**39**	138	10	2	**66**	86	3	1
14	102	9	1	**40**	153	2	2	**67**	158	10	1
15	108	4	2	**41**	183	9	1	**68**	167	5	1
16	180	20	1	**42**	195	2	1	**69**	112	7	1
17	176	9	1	**43**	173	14	1	**70**	137	8	1
18	124	6	1	**44**	148	19	2	**71**	39	7	1
19	88	3	1	**45**	176	31	2	**72**	66	5	1
20	242	19	1	**46**	148	22	1	**73**	66	8	1
21	181	14	1	**47**	150	11	2	**74**	159	6	1
22	146	9	1	**48**	186	10	1	**75**	244	0	0
23	213	11	1	**49**	102	5	2	**76**	124	4	1
24	114	4	1	**50**	109	17	2	**77**	93	0	0
25	156	6	1	**51**	233	30	1	**78**	32	0	0
26	171	11	1	**52**	118	16	1	***79***	2307	N/A	N/A
				53	212	41	1	**Total**	**13561**	**727**	**89**

Table C.5.1: Summary of AOs targeted to human *DMD* exons. Table shows the total number of AOs targeted to each exon, as recorded in our database, and the number of these AOs considered to be the best out of their set. The first and last exons of the gene cannot be skipped and are not AO targets, but they are included here for completeness.

Splice Factor	# of motifs	Avg. size	Splice Factor	# of motifs	Avg. size	Splice Factor	# of motifs	Avg. size
9G8	60	10.7	**hnRNP H1**	66	19.5	**QKI**	1	53
CUG-BP1	41	42.2	**hnRNP H2**	84	17.9	**RBM25**	2	7
DAZAP1	12	23.5	**hnRNP H3**	48	20.3	**RBM4**	8	7.4
ESRP1	2	24.5	**hnRNP I (PTB)**	105	27.7	**RBM5**	7	6.9
ESRP2	1	43	**hnRNP J**	1	7.0	**Sam68**	14	18.4
ETR3	29	10.8	**hnRNP K**	43	15.9	**SAP155**	1	19
FMRP	38	19.6	**hnRNP L**	143	23.3	**SC35**	171	9.9
Fox-1	4	6.0	**hnRNP LL**	12	28.9	**SF1**	17	7.0
Fox-2	3	6.0	**hnRNP M**	1	7	**SF2/ASF**	231	13.0
hnRNP A0	1	5.0	**hnRNP P (TLS)**	15	12.7	**SLM-1**	1	7.0
hnRNP A1	120	24.1	**hnRNP Q**	7	27.6	**SLM-2**	6	7.0
hnRNP A2/B1	38	20.0	**hnRNP U**	15	22.7	**SRm160**	1	34.0
hnRNP A3	2	28.0	**HTra2alpha**	7	18.9	**SRp20**	73	11.1
hnRNP C	13	19.2	**HTra2beta1**	22	15.0	**SRp30c**	23	11.1
hnRNP C1	9	11.2	**HuB**	44	8.1	**SRp38**	9	7.4
hnRNP C2	10	12.1	**HuC**	2	11.5	**SRp40**	68	15.6
hnRNP D	27	37.8	**HuD**	47	16.8	**SRp54**	1	6
hnRNP D0	1	6.0	**HuR**	67	31.9	**SRp55**	64	17.7
hnRNP DL	34	8.3	**KSRP**	21	20.4	**SRp75**	8	17.8
hnRNP E1	33	21.3	**MBNL1**	91	26.6	**TDP43**	17	18.8
hnRNP E2	36	17.6	**Nova-1**	25	10.8	**TIA-1**	36	17.5
hnRNP F	60	14.0	**Nova-2**	12	6.3	**TIAL1**	34	10.8
hnRNP G	1	50	**nPTB**	3	20.7	**YB-1**	28	22.8
			PSF	31	15.6	**ZRANB2**	19	6.3

Table C.5.2: Human splicing factors and their motifs. Table shows the names of the 71 splice factors considered in this experiment, the number of known RNA motifs for each, and the average size of those motifs.

C.5.1 Source Code (Python) For Measuring Frequencies of Splice Factor Motifs at Most Effective Exon-Skipping Antisense Oligomer Target Sites Versus All Other Sites

```
import csv
#import a module for printing to and reading from .csv files

def RevCom(seq):
    dictry = {'A':'T', 'T':'A', 'C':'G', 'G':'C', 'U':'A',
'a':'t', 't':'a', 'c':'g', 'g':'c', 'u':'a'}
    return "".join([dictry[base] for base in reversed(seq)])

#Convert data CSVs to lists

AOsfile = open('AntisenseOligomers.csv')
readAOs = csv.reader(AOsfile)
AOs = list(readAOs)
AOs.sort()
print("AOs loaded and sorted.")

exonsfile = open('Exons.csv')
readexons = csv.reader(exonsfile)
exons = list(readexons)
exons.sort()
print("Exons loaded and sorted.")

SFsfile = open('SpliceFactors.csv')
readSFs = csv.reader(SFsfile)
SFs = list(readSFs)
SFs.sort()
print('Splicing factors loaded and sorted.')


#Create a base-by-base "mask" of the AO coverage of each exon.
#"Y" if a base of the exon falls under at least one AO, "N" if
it doesn't.
AOmask = dict()
for exon in exons:
    mask = list()
    for n in exon[1]:
        mask.append(['N','N'])
    for ao in AOs:
        RCao = RevCom(ao[2])
        if ao[0] == exon[0]:
            frame = len(ao[2])
            compar = 0
            while (frame + compar) <= len(exon[1]):
                if (exon[1][compar:(compar+frame)]).upper() ==
RCao.upper():
                    rw = 0
                    while rw < len(ao[2]):
                        mask[(compar+rw)][0] = 'Y'
                        if ao[3] == 'Y':
```

```
                            mask[(compar+rw)][1] = 'Y'
                        rw+=1
                compar+=1
    #Send the results to a dictionary using exon numbers as
keys.
    #Each exon number therefore calls its AO mask as an array
    AOmask[exon[0]] = mask


#Create a base-by-base "mask" of the binding sites on each exon
for each SF.
#"Y" if a base of the exon is part of a site, "N" if it isn't.
SFmask = dict()
SFsubmask = dict()
for exon in exons:
    SFsubmask[exon[0]] = list('0'*len(exon[1]))
SFname = SFs[0][0]
for sf in SFs:
    if sf[0] != SFname:
        print(sf[0])
        SFmask[SFname] = dict()
        SFmask[SFname].update(SFsubmask)
        SFname = sf[0]
        for exon in exons:
            SFsubmask[exon[0]] = list('0'*len(exon[1]))
    perbase = (1/len(sf[1]))
    t4uSF = RevCom(RevCom(sf[1]))
    for exon in exons:
        frame = len(sf[1])
        compar = 0
        while (frame + compar) <= len(exon[1]):
            if (exon[1][compar:(compar+frame)]).upper() ==
t4uSF.upper():
                rw = 0
                while rw < len(sf[1]):
                    if perbase > float(SFsubmask[exon[0]]
[(compar+rw)]):
                        SFsubmask[exon[0]][(compar+rw)] =
perbase
                    rw+=1
            compar+=1
SFmask[SFname] = dict()
SFmask[SFname].update(SFsubmask)
    #Send the results to a dictionary using exon numbers as
keys.
    #Each exon number therefore calls its AO mask as an array

#Calculate frequencies for coincidence of best AOs and SF
motifs.
Results = list()
for sf in SFmask:
    bestsum = 0.0
    bestcount = 0
    restsum = 0.0
    restcount = 0
    for exon in SFmask[sf]:
```

```
        SFmask[sf][exon] = [float(i) for i in SFmask[sf][exon]]
        ziptest = list(zip(SFmask[sf][exon],AOmask[exon]))
        for posn in ziptest:
            if posn[1][0] == 'Y' and posn[1][1] == 'Y':
                if posn[0] != 0:
                    bestsum += posn[0]
                bestcount += 1
            if posn[1][0] == 'Y' and posn[1][1] == 'N':
                if posn[0] != 0:
                    restsum += posn[0]
                restcount += 1

Results.append([sf,str(bestsum),str(bestcount),str(restsum),str(
restcount)])


# Print results to a .csv file
with open ('output.csv', 'w', newline='') as csvfile:
    resultswriter = csv.writer(csvfile)
    for pair in Results:
        resultswriter.writerow(pair)
```

C.5.2 Guide to Formatting Input .csv Files

C.5.2.1 Formatting Antisense Oligomers Input File

Each row of `'AntisenseOligomers.csv'` should be ordered thusly:

```
[Target Exon],[AO Name],[AO sequence],[Y/N Most
Effective]
```

For example:

```
3,SampleAO1,ACUGAUCGUAGCUAGACUAG,N
3,SampleAO2,CUAGCUAGCGUAUCUGCAUG,Y
```

If appropriate, more than one AO per exon can be classed as “Most Effective”.

C.5.2.2 Formatting Exons Input File

Each row of `'Exons.csv'` should be ordered thusly:

```
[Exon number],[Exon sequence]
```

For example:

```
2,ttatatttaaagttgcttcctaacttttatttttttatttgcattttagA
TGAAAGAGAAGATGTTCAAAAGAAAACATTCACAAAATGGGTAAATGCACAAT
TTTCTAAGgtaagaatggtttgttactttacttttaagatctaagttgtgaaa
ttttc

3,atcattggaagtgtgctttgttaaattgagtgtatttttttttaatttcagT
TTGGGAAGCAGCATATTGAGAACCTCTTCAGTGACCTACAGGATGGGAGGCGC
CTCCTAGACCTCCTCGAAGGCCTGACAGGGCAAAAACTGgtatgtgacttatt
tttaagaaagttaactttaaacttagtagaatttca
```

Sequence can be upper-case, lower-case or a mix of both.

C.5.2.3 Formatting Splice Factors Input File

Each row of `'SpliceFactors.csv'` should be formatted thusly:

```
[Splice Factor Name],[Motif Sequence],[Motif Score]
```

For example:

```
9G8,UCGAGAGAU,8
9G8,UGGACAA,5
CUG-BP1,UGUGUGUGU,-8
```

Multiple motifs are permitted for each splice factor. Motif scores were not used in the calculations of this version of the script, as they would have hampered the assessment of overlapping sites, but they are retained as a dormant feature should they be required in future revisions.

C.5.3 Statistical Analysis

After processing, each row of the output results file `'output.csv'` will be formatted thusly:

```
[Splice Factor Name],[Number of motifs in most
effective AO target sites],[base coverage of most
effective AOs],[Number of motifs in other AO target
sites],[base coverage of other AOs]
```

For example:

```
Nova-2,  8.952380952,  2331,  23.3952381,   6414
SRp38,   4.0,          2331,  26.54761905,  6414
```

The Excel function BINOM.DIST can be used to determine whether a statistically significant relationship is present for each. For SRp38 in the above example, this would read as follows:

=BINOM.DIST(4.0,(4.0+26.54761905),(2331/(2331+6414)),FALSE)

This equates to a p-value of 0.0178, which is statistically significant.

Appendix D

Primer Sequences

Primer	Gene	Region	F/R	Sequence
696	DMD	e45	F	AATTGGGAAGCCTGAATCTG
1726	DMD	i44	R	CAATTTCGAAAAAACAAATCAAAG
4025	DMD	e44	R	CTGTTCAGCTTCTGTTAGCC
4047	DMD	e46	R	GCAATGTTATCTGCTTCCTCCAAC
4139	DMD	i44	R	CCTTCAGAACCTGATCTTTAAG
4142	DMD	i45	F	TTGTGTCCCAGTTTGCATTAAC
4145	DMD	i47	R	AACACATGTGACGGAAGAGATG
4206	DMD	e79	R	GCACACTTTAGTTTACAATC
4213	DMD	e1	F	TCAGTTACTGTGTTGACTCAC
4322	DMD	e47	F	GCTCCCATAAGCCCAGAAGAGC
4334	DMD	e18	R	CAGTTATATCAACATCCAACC
4336	DMD	e25	R	GTCTCAAGTCTCGAAGCAAAC
4349	DMD	e40	F	CTCTAGAAATTTCTCATC
4351	DMD	e40	F	GGTATCAGTACAAGAGGCAG
4353	DMD	e43	F	GCAACGCCTGTGGAAAGGGTG
4357	DMD	e51	F	CTAGAAATGCCATCTTCCTTG
4359	DMD	e51	F	CTGCTCTGGCAGATTTCAAC
4360	DMD	e58	R	CTCCTGGTAGAGTTTCTCTAG
4411	DMD	e53	F	GATTCAACACAATGGCTGGAA
4424	DMD	e55	R	CTGTAGGACATTGGCAGTTGTTTC
4456	DMD	e44	F	CTGTTGAGAAATGGCGGCGTT
4539	DMD	e43	R	GTAGCTTCACCCTTTCCACAG
4604	DMD	e52	R	TGCTGGTCTTGTTTTTCAAATTTTGGGC
4621	DMD	e33	R	CTGCTTTTTCTGTACAATCTG
4626	DMD	e48	F	TGTGGTTATCTCCTATTAGG
4633	DMD	e30	R	GAGCTGCGTCCACCTTGTCTGC
4647	DMD	e5	R	CCATCTACGATGTCAGTACTTCC
4677	DMD	e27	R	GCTATGACACTATTTACAGACTC
4683	DMD	e55	R	GAGTCTTCTAGGAGCCTTTCCTTA
4733	DMD	e42	F	CACACTGTCCGTGAAGAAACGATG
4834	DMD	e27	F	GTGAAACTCCTTACTGAGTCTG

Primer	Gene	Region	F/R	Sequence
4839	DMD	e56	R	GTGATAAACATCTGTGTGAGC
4846	DMD	e53	R	CATTCAACTGTTGCCTCCGG
4849	DMD	e53	R	CCAGCCATTGTGTTGAATCC
4851	DMD	e51	R	CAGTTTCCTTAGTAACCACAGG
4897	DMD	e1	F	TCCTGGCATCAGTTACTGTGTTG
4898	DMD	e1	F	CTCACTCAGTGTTGGGATCACTC
4903	DMD	e47	F	GCTCCCATAAGCCCAGAAGAGC
4904	DMD	e58	R	CTCCTGGTAGAGTTTCTCTAG
4963	DMD	e21	R	GGCCACAAAGTCTGCATCCAG
5229	DMD	e47	R	TTATCCACTGGAGATTTGTCTG
5306	DMD	e51	R	GGTAAGTTCTGTCCAAGCCCGG
5308	DMD	e10	R	CTCTCCATCAATGAACTGCC
5378	DMD	e17	R	CCGTAGTTACTGTTTCCATTA
5400	DMD	e45	R	CAGATTCAGGCTTCCCAATT
5401	DMD	e45	R	CCTGTAGAATACTGGCATCTGT
5413	DMD	e44	R	GCCGCCATTTCTCAACAGAT
5464	DMD	e50	R	ACCGCCTTCCACTCAGAGCTC
5652	DMD	e41	F	AGAGCAAATTTGCTCAGTTTCG
5659	DMD	e55	R	CAGGCAAGAAACTTTTCCAG
5677	DMD	e53	F	GAGGCAACAGTTGAATGAAATG
5742	DMD	i47	F	TCCTGGGTGCTTCATTGGTCG
5743	DMD	i47	R	GTCTTTGAATGAATAGTGCTGG
5839	DMD	e43	F	CAGGAAGCTCTCTCCCAGC
5867	DMD	i44	F	GGGGAATTTTCTTTCTTGGG
5868	DMD	i44	R	CGATGCAAAGTAGAGAGGGC
5869	DMD	i44	F	CAGGTAATGCAGACTGTGTG
5870	DMD	i44	R	CCAAGTACCAAAGCATCTCC
5871	DMD	i44	F	GCCTCATTTGTCTAGATAAG
5872	DMD	i44	R	CCAGTCACAATCTAGGTCTTC
5873	DMD	i44	F	CAAAATACTCCACTGCATAG
5874	DMD	i44	R	CTTTCTCTTTCAGTACCGCAG
5875	DMD	i44	F	CCTCCGTACATTGGTTTCTTAG

Primer	Gene	Region	F/R	Sequence
5876	DMD	i44	R	CACCACTGCTTATAACCTTGG
5877	DMD	i44	F	GATGACCTCATGACCTACTG
5878	DMD	i44	R	CTCTTAGGTCTCCCTCTTGTG
5879	DMD	i44	F	GTATCCGCTAGTTATTCCCC
5880	DMD	i44	R	GTCTCTCCACATAGTCACAC
5881	DMD	i46	F	CTGTCCTGGTTACCTACCTG
5882	DMD	i46	R	CACTGGAATAAGCTGGATGTC
5883	DMD	i46	F	GTGGGAATCATTTGAAGATGG
5884	DMD	i46	R	CAGGAGAAAATAGACTCATG
5885	DMD	i46	F	CTCTCAATGTCAGTTACCTC
5886	DMD	i46	R	CTGAAAACCTATCTCTCTATC
5887	DMD	i48	F	CCGGGACCAGTTTAGAATTTC
5888	DMD	i48	R	CTTTCTTGTCTGTTGTCTAGTG
5889	DMD	i48	F	GTACTAGCTCAAGTTCACCTG
5890	DMD	i48	R	GAGGAAGATTATGGTACCCAG
5891	DMD	i48	F	GATGGCTAGAACAGGGTTTG
5892	DMD	i48	R	GTGTCCAAGATGGGTAAAGAG
5893	DMD	i48	F	CATGAAACTACAAAGCCATGTG
5894	DMD	i48	R	CTGTTCAGTTTTGAGTGAAGC
5895	DMD	i48	F	GGTCTAGCAAACCTAATTGTAG
5896	DMD	i48	R	CGCATATATTGACTGCACTATC
5897	DMD	i48	F	CTTGAGCTTCCTTGCTTATCTG
5898	DMD	i48	R	GTGACTGAAGAGGATAGAGC
5899	DMD	i48	F	GTAGTAACTGTTGAAAGGTTGG
5900	DMD	i48	R	GCCTTGACCTCATTTTTGAAAG
5901	DMD	i49	F	GGGAGAAGCTGAACTTAAGGG
5902	DMD	i49	R	GTCCACAGCAAAAGTTGTGCG
5903	DMD	i49	F	CCTGCCATGTATTAGAACTG
5904	DMD	i49	R	CAGCTGGGAAAATGAATCATC
5905	DMD	i49	F	GAGTTTGGAGCTCTTGTGCG
5906	DMD	i49	R	CTGGCATTTGCCCATATTTTC
5907	DMD	i49	F	CTACAGTACACCAAACTCTTG
5908	DMD	i49	R	CCAGATAGGTAAATGAGGGG

Primer	Gene	Region	F/R	Sequence
5909	DMD	i49	F	CACTGGTGAGGGAATATGATC
5910	DMD	i49	R	CAGGTCCCTATAGGGACTAGG
5911	DMD	i49	F	CCAGGCTGAGTTAAGCTGTTC
5912	DMD	i49	R	CCAGACTTTGTGCCAGTTAC
5913	DMD	i50	F	GATTTACTAGCTCTGGGGTG
5914	DMD	i50	R	CATTGGAATCCCACTCCTTG
5915	DMD	i50	F	GATAGGCTTGATTCTCCCAC
5916	DMD	i50	R	GCTTTCCAGAAGAGGTGATG
5917	DMD	i50	F	CTTATCTCCACATTGACCTC
5918	DMD	i50	R	GAGGAGGCAGAGTATGTCCTC
5919	DMD	i50	F	CCAGTGTTTATTTGCCTGTC
5920	DMD	i50	R	CACTGAGGGTGAGATTACAG
5921	DMD	i50	F	GAGACTGGCTAGTTTGACAC
5922	DMD	i50	R	CACAGTCAGAACTAGTGTGC
5923	DMD	i50	F	CTGGCATATTCTGCTTACCC
5924	DMD	i50	R	GACCAGGTGAATTGTGGACT
5925	DMD	i50	F	CTTGTAACCCAAGATAGGGG
5926	DMD	i50	R	CTCCTGTTGTCATCCCTAAG
5927	DMD	i51	F	CCAGGTGTTCAGGATAATTC
5928	DMD	i51	R	CCAGGGCTATACTACACAAG
5929	DMD	i51	F	GCCGCTAAAGATGTACTAGG
5930	DMD	i51	R	GACATACGGTTTTGGAGGTG
5931	DMD	i51	F	GTATAGTGTCTCAAAGCCCC
5932	DMD	i51	R	CTCCCCATTGCTCCTTTATC
5933	DMD	i51	F	CTGTGATTTCCTCAGGGAGTC
5934	DMD	i51	R	GGGTTCGGTACTATGTGTCATC
5935	DMD	i51	F	CAGTTTATTGATGGCTGCCG
5936	DMD	i51	R	CGTATTGCTGCACTAGACCTG
5937	DMD	i51	F	CTTTGAGGGTCCAAGACTTC
5938	DMD	i51	R	CATCACCAAGAGTAGAGTGC
5939	DMD	i51	F	GGTCATGCACCCTACCTTAGTG
5940	DMD	i51	R	GGTTCCATGACACTACTACCTG
6015	DMD	i45	F	GTGAGACTTCATGTACGGTC

Primer	Gene	Region	F/R	Sequence
6016	DMD	i45	R	CTGTAAAGAGCCATTTGAGTC
6017	DMD	i45	F	GACCCCATACCAAATGAGAAG
6018	DMD	i45	R	GTCTCTAACTTGAGTCACCC
6019	DMD	i45	F	CCCAAAATCTACTGTGGGTG
6020	DMD	i45	R	GCACAGATCTGGGTCTAAATC
6021	DMD	i45	F	GTCCCAAATTCTATGATGCAG
6022	DMD	i45	R	CAAGCCTGTTGAGTGTATTGAG
6023	DMD	i45	F	CTACACTTCAATGCTTATCGC
6024	DMD	i45	R	GAGCATCTACCACATACCAG
6025	DMD	i45	F	GGCTCAGTTAGTTGATTGGATC
6026	DMD	i45	R	GAGATTACTAGATGTGGCCC
6027	DMD	i45	F	GGGTTTGAACCCAGTAATAC
6028	DMD	i45	R	CTAAATCAAACACTCCCCTC
6029	DMD	i47	F	CAGTTTGAAGCTGAGGGTTG
6030	DMD	i47	R	CAGATTTCTTGGATAGGCCG
6031	DMD	i47	F	GTTGTTCTGGCTCACTCATAC
6032	DMD	i47	R	GATTGCAAGGTGTGGTAGATTG
6033	DMD	i47	F	GGAAAGACTGCTCTGAAACTG
6034	DMD	i47	R	CAACCTGCCTACATTTGGAAAG
6035	DMD	i47	F	CGTCAGTTACATTTTGCAGC
6036	DMD	i47	R	CACCTAGCAACGTGCTGATC
6037	DMD	i47	F	CAGGGAAAGAAGTAGCTTGC
6038	DMD	i47	R	CTGCACATTGCCAATATCTC
6039	DMD	i47	F	GCCATTAAAGGGCTTGTTAC
6040	DMD	i47	R	GAACTCTGCCCTTATTACCAC
6041	DMD	i47	F	GTATGATGTGGAGAGTCAAG
6042	DMD	i47	R	GTGTGGGATTGAAGTGAGAAC
6043	DMD	i1	F	GTGACCTGGAAACTCACCATTC
6044	DMD	i1	F	CCTTTTACCTGACCACCCTG
6045	DMD	i11	F	GACTCTGGAAATAAGATGGC
6046	DMD	i11	F	GCAGTGAAGTTTGCGTATGTG
6047	DMD	i18	F	GGAAGAGGCTATCCTTCACCTG
6048	DMD	i18	F	CCATTCAGCTGTATATCCAG

Primer	Gene	Region	F/R	Sequence
6049	DMD	i25	F	GTATCACTCTGGCCATGTTCTG
6050	DMD	i25	F	GTAAAAGGAAGGAACAATGGCGC
6051	DMD	i25	F	GTAAAAGGAAGGAACAATGGC
6052	DMD	i25	F	CCAATTCAGCATTCCAGAGAG
6053	DMD	i26	F	CTGCAATTCCCAGATAGGTC
6054	DMD	i26	F	GGTAGAGCCCACAGGATTGAAG
6055	DMD	i45	F	GAGAAGACATACCAGTCGAGG
6056	DMD	i45	F	GTTTGAATGGGAGACGGAAGC
6057	DMD	i56	F	GTTGACTACAACAGTATGG
6058	DMD	i56	F	CCAGTGTGTATTTCATGCC
6059	DMD	i60	F	CAAGAAGCAGATGCCAGGGCAG
6060	DMD	i60	F	GCAGAATAACACCTGTGAAG
6061	DMD	i65	F	CTTCCTTCTACCTGACTGAAAAG
6062	DMD	i65	F	GATCTACCTTTTCTTAAGAGGGAG
6063	DMD	i77	F	GATTACTGTATGATCCTGGCC
6064	DMD	i77	F	CCTGCTCTGGAAAGTTGCTG
6187	NF1	i3	F	CTTGATTCTTACTTTTAAATC
6188	NF1	i3	F	CAGTTTAAGAAGAACGGAGC
6189	NF1	i6	F	GATGGTTTCCTTCATGGTTTG
6190	NF1	i6	F	CTTTGTAGAAGTATTGGACCG
6191	NF1	i6	F	GAGAATGTTCCATGTACAC
6192	NF1	i6	F	GAGAAGAATATATTGTGC
6193	NF1	i10a	F	GCATTCTGTAACTTAAAAGAG
6194	NF1	i10a	F	GAGTGCCGTCGAAGGAAAAAG
6195	NF1	i10b	F	GATGACATGTTTAACCTTTG
6196	NF1	i10b	F	GAGCTTCTTCAGTCCCTGGAG
6197	NF1	i19a	F	CCATGGATATAGAACTCAC
6198	NF1	i19a	F	CAGAGGGCCCATTGTACATG
6199	NF1	i23a	F	GCAAAGAAACTTAATTTCAAG
6200	NF1	i23a	F	CCTTAGGAAACGCTGACTGTC
6201	NF1	i30	F	GTAGGCCACATTGAGAAATC
6202	NF1	i30	F	CCTCTGGACCACTTACAAAG
6203	NF1	i30	F	CAAGAAAGGACTAAAATGGAG

Primer	Gene	Region	F/R	Sequence
6204	NF1	i30	F	GACTGTTCTTTCTTCGCCTC
6207	NF1	e8	R	GTGAGGGCTTATACGAAAGC
6208	NF1	e8	R	CTTGAGAATGGCTTACTTGG
6209	NF1	e13	R	CAAGAGGTTATGCACTGACAC
6210	NF1	e13	R	CACTTCTAGTTTGGTCTGGG
6211	NF1	e25	R	CAAAATCATTGAAAGGCCGC
6212	NF1	e25	R	GTGAAGAGAACATGATTGGC
6213	NF1	e33	R	CCTGTGGCTACTAAGAAAG
6214	NF1	e33	R	GTTGAAGGACAGCATCAGC
6217	DMD	e18	R	CTGGATATACAGCTGAATGG
6218	DMD	e18	R	GTAAGGAGATGTGGGAGTCAG
6219	DMD	i44	F	GAGGGAGCATAATGGACAAAG
6220	DMD	i44	R	GCAGATCTACAGTGGAATAG
6221	DMD	i44	F	CTCAATCCTTTCTAACATCCGC
6222	DMD	i44	R	CTTGCATAAAGTCTGACATCCC
6223	DMD	i44	F	GTTGTTGTCAGGAGTCACTTG
6224	DMD	i44	R	GACTGAGTCTACATTTCGGC
6225	DMD	i44	F	GTAGATCAAGGTTTTGCAGATC
6226	DMD	i44	R	CCAGAAATTGCCAATTTATCCC
6227	DMD	i44	F	CAAATGTGCAGAACACAACTC
6228	DMD	i44	R	GTCTGACAAAATCACATTCC
6229	DMD	i44	F	GGAGATGAATTGTTCTAGCCAC
6230	DMD	i44	R	CATACCCCATTCATCAGCAAAC
6231	DMD	i44	F	CATCCAAATGCCTGGGTGAG
6232	DMD	i44	R	GTTACAAGGAGATAAGCAGTG
6233	DMD	i44	F	CTTCAATCTTAGCGCTTGTCG
6234	DMD	i44	R	GAAGCTCTCAACCAGGAACTC
6235	DMD	i44	F	GCAATTTACAACTTGTGAGGCC
6236	DMD	i44	R	CATACAAAGACTCAGGATATGC
6237	DMD	i44	F	GGGAGCTGAGATGGAACATC
6238	DMD	i44	R	CTACTCTGAAGGGTCATTCG
6239	DMD	i44	F	GCACTCATTGCCTTACACTG
6240	DMD	i44	R	GAGGCATGAGAGGGTATAGTC

Primer	Gene	Region	F/R	Sequence
6241	DMD	i44	F	CCCAATTCATCCCATGTCTTTC
6242	DMD	i44	R	GTTTAATACTCGGCAAAGGAC
6243	DMD	i44	F	CCTCCTTGTCAATTTAGGGTG
6244	DMD	i44	R	CTGGTTGACTTGTCAGAGTGC
6245	DMD	i44	F	CAAAGTACACCTGTGAAGAGG
6246	DMD	i44	R	CTTTCCTCGACAGTGAATCTTC
6247	DMD	i44	F	GGCACGTTCATTTATCAGGTC
6248	DMD	i44	R	CGAGTGCTGTCAAATGCCTAC
6249	DMD	i44	F	GCACCATTCTCTTCCTTTAAC
6250	DMD	i44	R	CATTCACTTTCTTTGCAACTTC
6251	DMD	i44	F	CATGGTAATCTAGTGAGCCC
6252	DMD	i44	R	CCTTAGAGTGTCCGTTTCAT
6253	DMD	i44	F	CACAAGGGTGTTAAGAACTACC
6254	DMD	i44	R	GTAACACACGAAGAACCCTG
6255	DMD	i44	F	GTGGCTGTCCTACGAATAAG
6256	DMD	i44	R	GTGCTTGACATGATTTCCTG
6257	DMD	i44	F	GTTCTTGCGATAGTTTACTGAG
6258	DMD	i44	R	GGGCTCTAAATCCAAAGACTG
6259	DMD	i44	F	GGTGTATTTGCCTAGGGTGC
6260	DMD	i44	R	CTTAATGTCTTTGTCTGCTAGG
6261	DMD	i44	F	GGTGACTTTCAAATTTGCCTAC
6262	DMD	i44	R	CTTCATTGCAATCTTGTGCCC
6263	DMD	i44	F	CCCTACTACCTGGACATGAC
6264	DMD	i44	R	CTACTACAGGTTTTAAGGAGAC
6265	DMD	i44	F	CCATACCGTCCTCCATAATGG
6266	DMD	i44	R	CTACTTCCCCCATTCTTTGG
6267	DMD	i44	F	CTTTATCTGCAGGACAACTTG
6268	DMD	i44	R	CTACCTACTCATCTTTTCTGG
6269	DMD	i44	F	GAATTGTCTTTGACCAGCTTC
6270	DMD	i44	R	CCTTTGAGGTCTCCACTATGC
6271	DMD	i44	F	CTCTGCCAATAATGCTCTCAG
6272	DMD	i44	R	GTGAGGAAAGATGGTCGCTTG
6273	DMD	i44	F	GACCAGACTTGAAGGACTTAG

Primer	Gene	Region	F/R	Sequence
6274	DMD	i44	R	GTTACAAACATGTGCAGAGAG
6275	DMD	i47	F	GTTGTAAATATGGCTCTTGGG
6276	DMD	i47	R	GTTGGGCTTCTTAGGCAATAG
6277	DMD	i47	F	GGCCGGAAACCAAAGTCATC
6278	DMD	i47	R	CCAAGAAGTACTGTCCTATTC
6279	DMD	i47	F	CCCAGCGTCTAGTGAGGAAC
6280	DMD	i47	R	GAAGTTCTGCTCTACGTGATG
6281	DMD	i47	F	GAATGTAGCTGCATACTTAGC
6282	DMD	i47	R	CTTCACAGCAAGATAATGAGC
6283	DMD	i47	F	CATGTAGCCTAGGTGTGTAG
6284	DMD	i47	R	GGATTCCGAAGGCATAGTAC
6285	DMD	i47	F	CTCATACCGTCACACAGCTAG
6286	DMD	i47	R	GAAAATCAGTATCCCTGGAAC
6287	DMD	i49	F	GTCTCCTATTGCTACTTTCTG
6288	DMD	i49	R	GAAGTCATCACTCACGGAAG
6289	DMD	i49	F	CATCCATGGTATGTAGTTCAC
6290	DMD	i49	R	CAGTCTAGTGCTTCCTAATAAG
6291	DMD	i50	F	CAGCTCATCTTGTACACTGC
6292	DMD	i50	R	CAATTGGTGATTAGGGAGTTAC
6293	DMD	i50	F	CGGTGAAGCATAACTGAAGG
6294	DMD	i50	R	GAGATGCCATGAAACATTCC
6295	DMD	i50	F	CTTTCATCGGCAGTACAATCC
6296	DMD	i50	R	CTTTTCACAGGAGGCAGCTG
6297	DMD	i50	F	GTAGATTACAGTACATCTGGGC
6298	DMD	i50	R	GGTTAGTCCCCAACTTACTAAG
6299	DMD	i50	F	GCTGCCTGATTTATTGACCTC
6300	DMD	i50	R	CCTGTTAAACACTTGTACGC
6301	DMD	i50	F	CACATTGTCTTGGAAACATCAC
6302	DMD	i50	R	CTTATGTGCTAGGCAAGGTTC
6303	DMD	i51	F	CCTTACCAATGTTGCCTTTC
6304	DMD	i51	R	CATTTTAACGGCACAAAACAG
6305	DMD	i51	F	GATGTGTATGTCCCTCTGAG
6306	DMD	i51	R	GACCCTGGTAGGTACATCATG

Primer	Gene	Region	F/R	Sequence
6307	DMD	i51	F	GAACATGTGTTCTAGCAATTTC
6308	DMD	i51	R	GAATAAATGCTTGAGGGGTTG
6309	DMD	i51	F	CATATCCCTGTCCCTTTATCC
6310	DMD	i51	R	GGTAAGTTGGCCCCAAAGCC
6313	DMD	i52	F	GAGACATAAGGTGGAGCTAG
6314	DMD	i52	R	CAGTGTGTGTCTAAATGCTTC
6315	DMD	i52	F	CCAGGGAACTCCAGAATATG
6316	DMD	i52	R	CTCCTTTCAGCATAGGTTTG
6317	DMD	i52	F	GAAGGACTGCATAAAGAGG
6318	DMD	i52	R	CCTTCTCCCAATTCTATCAC
6319	DMD	i52	F	GTGGAGTCTATCGTGAAAGC
6320	DMD	i52	R	CAGAACTTGGTCTCAAGCAAG
6321	DMD	i52	F	CCTGTTTAAATGCCTTGTGG
6322	DMD	i52	R	CTGTTCTCCGAGGCTAAAAC
6323	DMD	i52	F	CTTAGCACTGCTTTTGCTAC
6324	DMD	i52	R	CGTGTATCTTTTCTGACCAC
6325	DMD	i52	F	CATAGTAGCACACAGTACAAG
6326	DMD	i52	R	CTTTGATAACGGAACACCCAC
6327	DMD	i44	F	GAAGACAGGATGTAGACAGTG
6328	DMD	i44	R	GGCATGGTATATTGCTCTGG
6329	DMD	i44	F	GCTGGTAGTGGCATGAGAG
6330	DMD	i44	R	GTTCTTTACCGTGGTTACTAC
6331	DMD	i44	F	GAGGTCGGTACACCATTGATG
6332	DMD	i44	R	GGACAATGAACACTAGCATAC
6333	DMD	i44	F	CTCATTCAGCAGTTTTCAGG
6334	DMD	i44	R	CAAGAGCTCAATATCCTGTG
6335	DMD	i44	F	GCATACATCCCCTTTCATGC
6336	DMD	i44	R	CGGTGAGAACACTTAAAATCG
6337	DMD	i44	F	GTAACACCTGACTCACAACG
6338	DMD	i44	R	GGAGACGGTAGGTATTAGAG
6339	DMD	i44	F	GTCTGCTGTGTCACTTATTG
6340	DMD	i44	R	CTTTGACGTTTACACATGCAAC
6341	DMD	i48	F	GTGAGCACAGAAGACTGCTC

Primer	Gene	Region	F/R	Sequence
6342	DMD	i48	F	CTTCAGGAAATGGAATGAAG
6343	DMD	i49	R	CACTTCAGAGTGTGGGCATC
6344	DMD	i49	R	GCTTCTTACAGAGTTCAGCTG
6345	DMD	i48	R	GAGCAGTCTTCTGTGCTCAC
6346	DMD	i48	R	CTTCATTCCATTTCCTGAAG
6347	DMD	i49	F	GATGCCCACACTCTGAAGTG
6348	DMD	i49	F	CAGCTGAACTCTGTAAGAAGC
6389	DMD	e52	F	GCAGGATTTGGAACAGAGGC
6390	DMD	e52	F	GGAAGAACTCATTACCGCTG
6391	DMD	e54	F	GCAGACAAATGTAGATGTGGC
6392	DMD	e54	F	CTTCTCCGGGATTATTCTGC
6397	NF1	e29	F	CTGGGACACTGCTCAATATC
6398	NF1	e29	F	GCCTAGTAGATGAGAACCAG
6411	DMD	i45	F	CTGTTATTGGGCTCTAAAGG
6412	DMD	i45	R	CCTAATTACAGCCAAATCCC
6413	DMD	i45	F	GCTGACAACATAATGAAGAGG
6414	DMD	i45	R	GGATTATGATTCAGAGGTTCC
6415	DMD	i44	F	CTCATTTAGAGCAGCAAGGTG
6416	DMD	i44	R	GTTAGGGATAGAGACGACAC
6417	DMD	i44	F	CACCTCTTCTCATCTAATTCC
6418	DMD	i44	R	CCTTATATGTCCCTCTTGCC
6419	DMD	i44	F	GGAACAGTATTCTAGGCAGG
6420	DMD	i44	R	CAACTCACTCATCCTCAATG
6421	DMD	i44	F	CATATGGTTTCTGGCCTTAG
6422	DMD	i44	R	CAACGGGAGTTATATCTGTC
6423	DMD	i44	F	GTTAGCTAAATCAAGGAAGGC
6424	DMD	i44	R	CACTTGTTAGTTTCTCTACTG
6425	DMD	i51	R	CCCAATTTGCATACTGGCTG
6426	DMD	i51	F	CAGTATCTCCATTTCACAGTG
6427	DMD	i51	R	CACTATCAATGGTTATTGGAG
6428	DMD	i51	F	CTTCCTCAGATGATTCACCAC
6429	DMD	i51	R	CAGGACCAGCTTCTTGAACG
6430	DMD	i51	F	CCTTGTGTAACTAACTGCAC

Primer	Gene	Region	F/R	Sequence
6431	DMD	i50	R	CCTACTTAGTAACGAAGCAGG
6432	DMD	i50	F	CAAAGGGGTCAGGGAAATATC
6433	DMD	i50	R	CACCCAATTACCCAATAGGATC
6434	DMD	i50	F	CTAAAGGTCCATGATATGTAG
6435	DMD	i50	R	CGATCACAATCTTCTGTGAAG
6436	DMD	i50	F	GACTGTATGGGTTAACTAGAG
6437	DMD	i47	R	CAAGCCACGTATTCCAGTTC
6438	DMD	i47	F	GCTCTGAGTTGCCAGGATG
6439	DMD	i47	R	CTGTGACTTAGGCAAAATGAC
6440	DMD	i47	F	GCCTATGGTAAGATTGGTTTC
6441	DMD	i47	R	GTGAGTGGAGGTGAGCTTC
6442	DMD	i47	F	CCCAAGTATAACTTTCCCAC
6443	DMD	i47	R	GATTTGTTACATAACACTATTGC
6444	DMD	i47	F	GCCTCACTCCCATTTCTGGC
6445	DMD	i50	R	CAAGTGATCCTGTCACCCAC
6446	DMD	i50	F	CATGTGGCATTCATATGTCTG
6447	DMD	i50	R	GGAAATCATTTATGCCTATCC
6448	DMD	i50	F	CTTATGGATACTTCAGTGCC
6449	DMD	i50	R	CCCTTGAGAAATATCTCCAAC
6450	DMD	i50	F	GGCACTTAGATGATGCCAAC
6451	DMD	i47	R	GTTTTGTGTGCCATGACATG
6452	DMD	i47	F	CAAAAGTGTTTATCCCTGTAG
6453	DMD	i47	R	CATCCCTCCCTTCTATGAAC
6454	DMD	i47	F	CTCCAGACTCCATGACTTTG
6455	DMD	i47	R	GAAAGGGAAAGGAACATTCC
6456	DMD	i47	F	GCATAAGTTGGGAACAGAGC
6457	DMD	i49	R	TGCTGTGAACTACAAAGCAC
6458	DMD	i49	F	GTTTTGCTTCATTGAGCAGC
6459	DMD	i49	R	CTCAGAATGAGGATAAAGAG
6460	DMD	i49	F	GAACCTAAGTCAACTGGTAG
6461	DMD	i47	R	GATAGTTTCAATAATATGACCATG
6462	DMD	i47	F	GATGAAGGCAAAGGAATAGG
6463	DMD	i47	R	CGTATCCTCAAATTCAGAAG

Primer	Gene	Region	F/R	Sequence
6464	DMD	i47	F	GATTATAGGCAGGCACCACC
6465	DMD	i47	R	CATACTGTTTAAACCCTGAGG
6466	DMD	i47	F	GTTTAGAGTACGGCTGGGAG
6467	DMD	i50	R	CAGGGTAACATAAATATGCC
6468	DMD	i50	F	CTCCTATTTCAGCAAGTATC
6469	DMD	i50	R	CATCCTCTTGGTTCAGGAAG
6470	DMD	i50	F	CAGAATGGTTTCATAAGCTG
6471	DMD	i50	R	GAAGTATCACTTGCCTTGAC
6472	DMD	i50	F	GCAATCATATCTTGGCCTCC
6473	DMD	i51	R	GCGACAGAGACTCCATCTC
6474	DMD	i51	F	CAGTCATCTCCAACATGAGAC
6475	DMD	i51	R	GTAGAAACTCTTCTGTGTGG
6476	DMD	i51	F	GTTTGTAAGGAAAACCAGTGC
6477	DMD	i51	R	GAATTTGTCAGACTATGTAAGC
6478	DMD	i51	F	GTTGGAGAAAGAGAAAGCAG
6594	DMD	i45	R	GCTTCATAGTCATATTCATAGC
6595	DMD	i45	F	GTAGGAAAGACAGATCTTTGC
6596	DMD	i45	F	CATCATCCACATAATAGGTGG
6597	DMD	i45	R	CAATGAGTTGTTCTAGCTTCC
6598	DMD	i45	R	CCACTTTATGAATTCTCTCCG
6599	DMD	i45	F	CTATGAACAGGTATAAACCTG
6600	DMD	i51	R	GTTAACTACAAGCCCTAGGC
6601	DMD	i51	F	GTCCTACAAATCTGTACAGC
6602	DMD	i51	R	GTATACATGTAGGACAGTGG
6603	DMD	i51	F	CACGATGGCAAGTATTTGTG

Appendix E

Reagents and Their Suppliers

Reagent / chemical	Supplier
100bp molecular size marker	Geneworks
Acrylamide/Bis Solution (29:1)	Bio-Rad Laboratories
Agarose powder	Scientifix
Ammonium persulfate (APS: (NH2)2S3O8)	Sigma-Aldrich
AmpliTaq Gold DNA polymerase and reaction buffer	Applied Biosystems
Baxter sterile water	Baxter Healthcare
Beta Tubulin monoclonal antibody (βTUB)	Thermo Scientific
Bromophenol blue (BPB)	Sigma-Aldrich
Chick embryo extract (CEE)	Jomar Diagnostics
Deoxynucleotide triphosphates (dNTPs)	Life Technologies
Diffinity Rapid tips	Sigma-Aldrich
Dimethyl sulfoxide (DMSO: C2H6OS)	Sigma-Aldrich
Dithiothreitol (DTT)	Roche Diagnostics
Dulbecco's modified Eagle's Medium (DMEM)	Life Technologies
Ethanol (C2H5OH)	Merck
Fetal calf serum	Scientifix
Gel loading buffer	Made in house
Glycerol (C3H8O3)	Sigma-Aldrich
Glycine	Sigma-Aldrich
HAMs-F10 medium	Life Technologies
Horse serum	Life Technologies
MagicMark XP Western protein standard	Life Technologies
Matrigel	Becton Dickinson Life Technologies
NCL-DYS2 Dystrophin monoclonal antibody	Novocastra
Phenylmethylsulphonyl fluoride (PMSF)	Aldrich
Phosphate buffer saline (PBS)	Made in house
Poly-D-Lysine	Sigma-Aldrich
Precision Plus Protein Kaleidoscope Standards	Bio-Rad Laboratories
Protease inhibitor cocktail (P8340)	Sigma-Aldrich
Polyvinylidene Fluoride (PVDF) transfer membrane	Pall
RedSafe nucleic acid stain	iNtRON Biotechnologies
Sodium dodecyl sulfate (SDS: C12H25O4SNa)	Sigma-Aldrich
Superscript III one-step RT-PCR with Platinum Taq	Life Technologies
TEMED (N,N,N',N'- tetramethylethylenediamene: C6H16N2)	Sigma-Aldrich
Tissue culture grade water	Sigma-Aldrich
Tris-acetate-EDTA (TAE) buffer	Made in house
Tris-Glycine-SDS western transfer buffer (WTB)	Made in house
Trizma base (C4H11NO3)	Sigma-Aldrich
Trypan Blue (C34H24N6O14S4Na4)	Sigma-Aldrich
Trypsin (diluted 0.25% in PBS)	Life Technologies
Western Breeze Chemiluminescent immunodetection kit	Life Technologies
Tris Buffer (pH 8.3)	Made in house

Glossary

ADV: Adenovirus

AO: Antisense oligonucleotide

Array-CGH: Array-comparative genomic hybridisation

BMD: Becker Muscular Dystrophy

cDNA: Complementary DNA

CUG-BP: CUG-binding protein

ddH2O: Double-distilled H2O

DM1: Dystrophia Myotonica Type 1

DM2: Dystrophia Myotonica Type 2

DMD: Duchenne Muscular Dystrophy

DMPK: Dystrophia Myotonica Protein Kinase.

DMSO: Dimethyl sulfoxide

DNA: Deoxyribonucleic acid

dNTP: Deoxynucleotide triphosphate

EDTA: Ethylenediaminetetraacetic acid

ESE: Exonic silencer element

FOX2: Forkhead box protein P2.

gDNA: Genomic DNA

hnRNP: Heterogenous ribonucleoprotein particle

LOVD: Leiden Online Variation Database

MBNL1: Muscleblind-like protein

MLPA: Multiplex ligation-dependent probe amplification

mRNA: Messenger RNA

NCBI: National Centre for Biotechnology Information

NF1: Neurofibromin 1

NGS: Next generation sequencing

nNOS: Neuronal nitric oxide synthase

NMD: Nonsense mediated decay

PBS: Phosphate buffered saline

PCR: Polymerase chain reaction

PE: Pseudoexon

PEDSS: Pseudoexon donor splice site

RBM5: RNA binding motif protein 5

RNA: Ribonucleic acid

RT-PCR: Reverse transcription polymerase chain reaction

SF: Splice/splicing factor

SMN2: Survival of motor neuron 2, centromeric.

SNP: Single nucleotide polymorphism

snRNP: Small nuclear ribonucleo proteins

SRp: Serine and arginine rich protein

TAE: Tris-acetate-EDTA

WANRI: Western Australian Neuroscience Research Institute

WGS: Whole genome sequencing

WTS: Whole transcriptome sequencing

XLDC: X-Linked Dilated Cardiomyopathy

ZNF9: Zinc finger protein 9.

ZRANB2: Zinc finger Ran-binding domain-containing protein 2

References

Aartsma-Rus, A., J. C. Van Deutekom, I. F. Fokkema, G. J. Van Ommen and J. T. Den Dunnen (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.

Aartsma-Rus, A., W. E. Kaman, R. Weij, J. T. den Dunnen, G.-J. B. van Ommen and J. C. T. van Deutekom (2006). "Exploring the Frontiers of Therapeutic Exon Skipping for Duchenne Muscular Dystrophy by Double Targeting within One or Multiple Exons." Mol Ther **14**(3): 401-407.

Aartsma-Rus, A., L. van Vliet, M. Hirschi, A. A. Janson, H. Heemskerk, C. L. de Winter, S. de Kimpe, J. C. van Deutekom, P. A. t Hoen and G. J. van Ommen (2009). "Guidelines for antisense oligonucleotide design and insight into splice-modulating mechanisms." Mol Ther **17**(3): 548-553.

Aartsma-Rus, A. and G. J. van Ommen (2010). "Progress in therapeutic antisense applications for neuromuscular disorders." Eur J Hum Genet **18**(2): 146-153.

Abdel-Salam, E., I. Abdel-Meguid and S. S. Korraa (2009). "Markers of degeneration and regeneration in Duchenne muscular dystrophy." Acta Myol **28**(3): 94-100.

Adams, A. M., P. L. Harding, P. L. Iversen, C. Coleman, S. Fletcher and S. D. Wilton (2007). "Antisense oligonucleotide induced exon skipping and the dystrophin gene transcript: cocktails and chemistries." BMC Mol Biol **8**: 57.

Ahn, A. H. and L. M. Kunkel (1993). "The Structural and Functional Diversity of Dystrophin." Nature Genetics **3**(4): 283-291

Allen, D. G. and N. P. Whitehead (2011). "Duchenne muscular dystrophy – What causes the increased membrane permeability in skeletal muscle?" The International Journal of Biochemistry & Cell Biology **43**(3): 290-294.

Amann, K. J., B. A. Renley and J. M. Ervasti (1998). "A Cluster of Basic Repeats in the Dystrophin Rod Domain Binds F-actin through an Electrostatic Interaction." Journal of Biological Chemistry **273**(43): 28419-28423.

Ameur, A., A. Zaghlool, J. Halvardson, A. Wetterbom, U. Gyllensten, L. Cavelier and L. Feuk (2011). "Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain." Nat Struct Mol Biol **18**(12): 1435-1440.

Anthony, K., S. Cirak, S. Torelli, G. Tasca, L. Feng, V. Arechavala-Gomeza, A. Armaroli, M. Guglieri, C. S. Straathof, J. J. Verschuuren, A. Aartsma-Rus, P. Helderman-van den Enden, K. Bushby, V. Straub, C. Sewry, A. Ferlini, E. Ricci, J. E. Morgan and F. Muntoni (2011). "Dystrophin quantification and clinical correlations in Becker muscular dystrophy: implications for clinical trials." Brain **134**(Pt 12): 3547-3559.

Antonini, G., F. Soscia, F. Giubilei, A. De Carolis, F. Gragnani, S. Morino, A. Ruberto and R. Tatarelli (2006). "Health-related quality of life in myotonic dystrophy type 1 and its relationship with cognitive and emotional functioning." J Rehabil Med **38**(3): 181-185.

Aslanidis, C., G. Jansen, C. Amemiya, G. Shutler, M. Mahadevan, C. Tsilfidis, C. Chen, J. Alleman, N. G. Wormskamp, M. Vooijs and et al. (1992). "Cloning of the essential myotonic dystrophy region and mapping of the putative defect." Nature **355**(6360): 548-551.

Atkinson, G. C., S. L. Baldauf and V. Hauryliuk (2008). "Evolution of nonstop, no-go and nonsense-mediated mRNA decay and their termination factor-derived components." BMC Evol Biol **8**: 290.

Bacolla, A. and R. D. Wells (2004). "Non-B DNA conformations, genomic rearrangements, and human disease." J Biol Chem **279**(46): 47411-47414.

Banks, G. B., P. Gregorevic, J. M. Allen, E. E. Finn and J. S. Chamberlain (2007). "Functional capacity of dystrophins carrying deletions in the N-terminal actin-binding domain." Hum Mol Genet **16**(17): 2105-2113.

Barbaro, M., S. Bens, A. Haake, M. Peter, J. Bramswig, P. M. Holterhus, J. P. Lopez-Siguero, U. Menken, M. Mix, W. G. Sippell, A. Wedell and F. G. Riepe (2012). "Multiplex ligation-dependent probe amplification analysis of the NR0B1(DAX1) locus enables explanation of phenotypic differences in patients with X-linked congenital adrenal hypoplasia." Horm Res Paediatr **77**(2): 100-107.

Beggs, A. H., M. Koenig, F. M. Boyce and L. M. Kunkel (1990). "Detection of 98% of DMD/BMD gene deletions by polymerase chain reaction." Hum Genet **86**(1): 45-48.

Beggs, A. H., E. P. Hoffman, J. R. Snyder, K. Arahata, L. Specht, F. Shapiro, C. Angelini, H. Sugita and L. M. Kunkel (1991). "Exploring the molecular basis for variability among patients with Becker muscular dystrophy: dystrophin gene and protein studies." Am J Hum Genet **49**(1): 54-67.

Berget, S. M. (1995). "Exon recognition in vertebrate splicing." J Biol Chem **270**(6): 2411-2414.

Beroud, C., A. Carrie, C. Beldjord, N. Deburgrave, S. Llense, N. Carelle, C. Peccate, J. M. Cuisset, F. Pandit, F. Carre-Pigeon, M. Mayer, R. Bellance, D. Recan, J. Chelly, J. C. Kaplan and F. Leturcq (2004). "Dystrophinopathy caused by mid-intronic substitutions activating cryptic exons in the DMD gene." Neuromuscul Disord **14**(1): 10-18.

Bettencourt, C., M. Raposo, R. Ros, R. Montiel, J. Bruges-Armas and M. Lima (2013). "Transcript Diversity of Machado-Joseph Disease Gene (ATXN3) Is Not Directly Determined by SNPs in Exonic or Flanking Intronic Regions." Journal of Molecular Neuroscience **49**(3): 539-543.

Bies, R. D., C. T. Caskey and R. Fenwick (1992). "An intact cysteine-rich domain is required for dystrophin function." Journal of Clinical Investigation **90**(2): 666-672.

Bo, X., S. Lou, D. Sun, W. Shu, J. Yang and S. Wang (2006). "Selection of antisense oligonucleotides based on multiple predicted target mRNA structures." BMC Bioinformatics **7**: 122.

Bos, D. H., C. Mayfield and D. J. Minchella (2009). "Analysis of regulatory protease sequences identified through bioinformatic data mining of the Schistosoma mansoni genome." Bmc Genomics **10**.

Buratti, E. and F. E. Baralle (2004). "Influence of RNA secondary structure on the pre-mRNA splicing process." Mol Cell Biol **24**(24): 10505-10514.

Cagliani, R., M. Sironi, E. Ciafaloni, A. Bardoni, F. Fortunato, A. Prelle, M. Serafini, N. Bresolin and G. P. Comi (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.

Camps-Valls, G., A. M. Chalk, A. J. Serrano-Lopez, J. D. Martin-Guerrero and E. L. Sonnhammer (2004). "Profiled support vector machines for antisense oligonucleotide efficacy prediction." BMC Bioinformatics **5**: 135.

Cartegni, L., J. Wang, Z. Zhu, M. Q. Zhang and A. R. Krainer (2003). "ESEfinder: A web resource to identify exonic splicing enhancers." Nucleic Acids Res **31**(13): 3568-3571.

Cavalcanti, G. M., A. D. B. Oliveira, T. D. Assis, L. M. C. Chimelli, P. L. de Medeiros and D. L. da Mota (2011). "Histochemistry and Morphometric Analysis of Muscle Fibers from Patients with Duchenne Muscular Dystrophy (DMD)." International Journal of Morphology **29**(3): 934-938.

Chen, M. and J. L. Manley (2009). "Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches." Nat Rev Mol Cell Biol **10**(11): 741-754.

Craig, R. and L. Liao (2006). "Prediction of antisense oligonucleotide efficacy using local and global structure information with support vector machines." ICMLA 2006: 5th International Conference on Machine Learning and Applications, Proceedings: 199-204.

Danglot, G., V. Regnier, D. Fauvet, G. Vassal, M. Kujas and A. Bernheim (1995). "Neurofibromatosis 1 (NF1) mRNAs expressed in the central nervous system are differentially spliced in the 5' part of the gene." Hum Mol Genet **4**(5): 915-920.

Darras B.T., Miller D.T., Urion D.K.. “Dystrophinopathies.” 2000 Sep 5 [Updated 2014 Nov 26]. In: Pagon R.A., Adam M.P., Ardinger H.H., et al. (editors). GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2015.

Deburgrave, N., F. Daoud, S. Llense, J. C. Barbot, D. Recan, C. Peccate, A. H. Burghes, C. Beroud, L. Garcia, J. C. Kaplan, J. Chelly and F. Leturcq (2007). "Protein- and mRNA-based phenotype-genotype correlations in DMD/BMD with point mutations and molecular basis for BMD with nonsense and frameshift mutations in the DMD gene." Hum Mutat **28**(2): 183-195.

Desguerre, I., C. Christov, M. Mayer, R. Zeller, H. M. Becane, S. Bastuji-Garin, F. Leturcq, C. Chiron, J. Chelly and R. K. Gherardi (2009). "Clinical heterogeneity of duchenne muscular dystrophy (DMD): definition of sub-phenotypes and predictive criteria by long-term follow-up." PLoS One **4**(2): e4347.

Desmet, F. O., D. Hamroun, M. Lalande, G. Collod-Beroud, M. Claustres and C. Beroud (2009). "Human Splicing Finder: an online bioinformatics tool to predict splicing signals." Nucleic Acids Res **37**(9): e67.

Dwi Pramono, Z. A., Y. Takeshima, A. Surono, T. Ishida and M. Matsuo (2000). "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." Biochem Biophys Res Commun **267**(1): 321-328.

Echigoya, Y., V. Mouly, L. Garcia, T. Yokota and W. Duddy (2015). "In silico screening based on predictive algorithms as a design tool for exon skipping oligonucleotides in Duchenne muscular dystrophy." PLoS One **10**(3): e0120058.

Emery, A. E. (2002). "The muscular dystrophies." Lancet **359**(9307): 687-695.

Feener, C. A., M. Koenig and L. M. Kunkel (1989). "Alternative splicing of human dystrophin mRNA generates isoforms at the carboxy terminus." Nature **338**(6215): 509-511

Feng, J., J. Y. Yan, C. H. Buzin, S. S. Sommer and J. A. Towbin (2002). "Comprehensive mutation scanning of the dystrophin gene in patients with nonsyndromic X-linked dilated cardiomyopathy." J Am Coll Cardiol **40**(6): 1120-1124.

Ferlini, A., N. Galie, L. Merlini, C. Sewry, A. Branzi and F. Muntoni (1998). "A novel Alu-like element rearranged in the dystrophin gene causes a splicing mutation in a family with X-linked dilated cardiomyopathy." Am J Hum Genet **63**(2): 436-446.

Fernandez-Rodriguez, J., J. Castellsague, L. Benito, Y. Benavente, G. Capella, I. Blanco, E. Serra and C. Lazaro (2011). "A mild neurofibromatosis type 1 phenotype produced by the combination of the benign nature of a leaky NF1-splice mutation and the presence of a complex mosaicism." Hum Mutat **32**(7): 705-709.

Ferreiro, V., F. Giliberto, G. M. Muniz, L. Francipane, D. M. Marzese, A. Mampel, M. Roque, G. D. Frechtel and I. Szijan (2009). "Asymptomatic Becker muscular dystrophy in a family with a multiexon deletion." Muscle Nerve **39**(2): 239-243.

Finsterer, J., C. Stöllberger, M. Gencik, R. Höftberger, J. Rahimi and J. Mokocki (2015). "Syncope and hyperCKemia as minimal manifestations of short CTG repeat expansions in myotonic dystrophy type 1." Revista Portuguesa de Cardiologia **34**(5): 361.e361-361.e364.

Fletcher, S., A. M. Adams, R. D. Johnsen, K. Greer, H. M. Moulton and S. D. Wilton (2010). "Dystrophin isoform induction in vivo by antisense-mediated alternative splicing." Mol Ther **18**(6): 1218-1223.

Fletcher, S., P. L. Meloni, R. D. Johnsen, B. L. Wong, F. Muntoni and S. D. Wilton (2013). "Antisense suppression of donor splice site mutations in the dystrophin gene transcript." Mol Genet Genomic Med **1**(3): 162-173.

Foley, S. B., J. J. Rios, V. E. Mgbemena, L. S. Robinson, H. L. Hampel, A. E. Toland, L. Durham and T. S. Ross (2015). "Use of Whole Genome Sequencing for Diagnosis and Discovery in the Cancer Genetics Clinic." Ebiomedicine **2**(1): 74-81.

Fu, X. D. and T. Maniatis (1992). "The 35-Kda Mammalian Splicing Factor Sc35 Mediates Specific Interactions between U1 and U2 Small Nuclear Ribonucleoprotein-Particles at the 3' Splice Site." Proceedings of the National Academy of Sciences of the United States of America **89**(5): 1725-1729.

Gazzoli, I., I. Pulyakhina, N. E. Verwey, Y. Ariyurek, J. F. Laros, P. A. t Hoen and A. Aartsma-Rus (2016). "Non-sequential and multi-step splicing of the dystrophin transcript." RNA Biol **13**(3): 290-305.

Greer, K., K. Mizzi, E. Rice, L. Kuster, R. A. Barrero, M. I. Bellgard, B. J. Lynch, A. R. Foley, O. R. E, S. D. Wilton and S. Fletcher (2015). "Pseudoexon activation increases phenotype severity in a Becker muscular dystrophy patient." Mol Genet Genomic Med **3**(4): 320-326.

Gualandi, F., P. Rimessi, C. Trabanelli, P. Spitali, M. Neri, T. Patarnello, C. Angelini, S. C. Yau, S. Abbs, F. Muntoni, E. Calzolari and A. Ferlini (2006). "Intronic breakpoint definition and transcription analysis in DMD/BMD patients with deletion/duplication at the 5 ' mutation hot spot of the dystrophin gene." Gene **370**: 26-33.

Gurvich, O. L., T. M. Tuohy, M. T. Howard, R. S. Finkel, L. Medne, C. B. Anderson, R. B. Weiss, S. D. Wilton and K. M. Flanigan (2008). "DMD pseudoexon mutations: splicing efficiency, phenotype, and potential therapy." Ann Neurol **63**(1): 81-89.

Halvardson, J., A. Zaghlool and L. Feuk (2013). "Exome RNA sequencing reveals rare and novel alternative transcripts." Nucleic Acids Res **41**(1): e6.

Hamed, S. A., A. J. Sutherland-Smith, J. R. M. Gorospe, J. Kendrick-Jones and E. P. Hoffman (2005). "DNA sequence analysis for structure/function and mutation studies in Becker muscular dystrophy." Clinical Genetics **68**(1): 69-79.

Harding, P. L., A. M. Fall, K. Honeyman, S. Fletcher and S. D. Wilton (2007). "The influence of antisense oligonucleotide length on dystrophin exon skipping." Mol Ther **15**(1): 157-166.

Hastings, P. J., G. Ira and J. R. Lupski (2009). "A microhomology-mediated break-induced replication model for the origin of human copy number variation." PLoS Genet **5**(1): e1000327.

Heemskerk, H. A., C. L. de Winter, S. J. de Kimpe, P. van Kuik-Romeijn, N. Heuvelmans, G. J. Platenburg, G. J. van Ommen, J. C. van Deutekom and A. Aartsma-Rus (2009). "In vivo comparison of 2'-O-methyl phosphorothioate and morpholino antisense oligonucleotides for Duchenne muscular dystrophy exon skipping." J Gene Med **11**(3): 257-266.

Ho, G., M. Cardamone and M. Farrar (2015). "Congenital and childhood myotonic dystrophy: Current aspects of disease and future directions." World J Clin Pediatr **4**(4): 66-80.

Hoffman, E. P., R. H. Brown, Jr. and L. M. Kunkel (1987). "Dystrophin: the protein product of the Duchenne muscular dystrophy locus." Cell **51**(6): 919-928.

Hoffman, E. P., K. H. Fischbeck, R. H. Brown, M. Johnson, R. Medori, J. D. Loike, J. B. Harris, R. Waterston, M. Brooke, L. Specht and et al. (1988). "Characterization of dystrophin in muscle-biopsy specimens from patients with Duchenne's or Becker's muscular dystrophy." N Engl J Med **318**(21): 1363-1368.

Hong, X., D. G. Scofield and M. Lynch (2006). "Intron size, abundance, and distribution within untranslated regions of genes." Molecular Biology and Evolution **23**(12): 2392-2404.

Howard, J. M. and J. R. Sanford (2015). "The RNAissance family: SR proteins as multifaceted regulators of gene expression." Wiley Interdisciplinary Reviews-Rna **6**(1): 93-110.

Hsu, S. N. and K. J. Hertel (2009). "Spliceosomes walk the line Splicing errors and their impact on cellular function." Rna Biology **6**(5): 526-530.

Hug, N., D. Longman and J. F. Caceres (2016). "Mechanism and regulation of the nonsense-mediated decay pathway." Nucleic Acids Res **44**(4): 1483-1495.

Inagaki, H., T. Ohye, H. Kogo, T. Kato, H. Bolor, M. Taniguchi, T. H. Shaikh, B. S. Emanuel and H. Kurahashi (2009). "Chromosomal instability mediated by non-B DNA: cruciform conformation and not DNA sequence is responsible for recurrent translocation in humans." Genome Res **19**(2): 191-198.

Ishibashi, K., Y. Takeshima, M. Yagi, A. Nishiyama and M. Matsuo (2006). "Novel cryptic exons identified in introns 2 and 3 of the human dystrophin gene with duplication of exons 8-11." Kobe J Med Sci **52**(3-4): 61-75.

Ishikawa-Sakurai, M., M. Yoshida, M. Imamura, K. E. Davies and E. Ozawa (2004). "ZZ domain is essentially required for the physiological binding of dystrophin and utrophin to beta-dystroglycan." Human Molecular Genetics **13**(7): 693-702.

Ishmukhametova, A., P. K. Van Kien, D. Mechin, D. Thorel, M. C. Vincent, F. Rivier, C. Coubes, V. Humbertclaude, M. Claustres and S. Tuffery-Giraud (2012). "Comprehensive oligonucleotide array-comparative genomic hybridization analysis: new insights into the molecular pathology of the DMD gene." European Journal of Human Genetics **20**(10): 1096-1100.

Ikezawa, M., N. Minami, M. Takahashi, Y. Goto, T. Miike and I. Nonaka (1998). "Dystrophin gene analysis on 130 patients with Duchenne muscular dystrophy with a special reference to muscle mRNA analysis." Brain Dev **20**(3): 165-168.

Jin, W., Z. Niu, D. Xu and X. Li (2012). "RBM5 promotes exon 4 skipping of AID pre-mRNA by competing with the binding of U2AF65 to the polypyrimidine tract." FEBS Lett **586**(21): 3852-3857.

Jog, S. P., S. Paul, W. Dansithong, S. Tring, L. Comai and S. Reddy (2012). "RNA Splicing Is Responsive to MBNL1 Dose." Plos One **7**(11).

Juan-Mateu, J., L. Gonzalez-Quereda, M. J. Rodriguez, E. Verdura, K. Lazaro, C. Jou, A. Nascimento, C. Jimenez-Mallebrera, J. Colomer, S. Monges, F. Lubieniecki, M. E. Foncuberta, S. I. Pascual-Pascual, J. Molano, M. Baiget and P. Gallano (2013). "Interplay between DMD point mutations and splicing signals in Dystrophinopathy phenotypes." PLoS One **8**(3): e59916.

Kalsotra, A., X. Xiao, A. J. Ward, J. C. Castle, J. M. Johnson, C. B. Burge and T. A. Cooper (2008). "A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart." Proc Natl Acad Sci U S A **105**(51): 20333-20338.

Kannu, P., M. Nour, M. Irving, J. Xie, D. Loder, J. Lai, O. Islam, J. MacKenzie and L. Messiaen (2013). "Paraspinal ganglioneuroma in the proband of a large family with mild cutaneous manifestations of NF1, carrying a deep NF1 intronic mutation." Clin Genet **83**(2): 191-194.

Kervestin, S. and A. Jacobson (2012). "NMD: a multifaceted response to premature translational termination." Nat Rev Mol Cell Biol **13**(11): 700-712.

Kesari, A., L. N. Pirra, L. Bremadesam, O. McIntyre, E. Gordon, A. L. Dubrovsky, V. Viswanathan and E. P. Hoffman (2008). "Integrated DNA, cDNA, and protein studies in Becker muscular dystrophy show high exception to the reading frame rule." Human mutation **29**(5): 728-737.

Khalighi, K., A. Kodali, S. B. Thapamagar and S. R. Walker (2015). "Cardiac involvement in myotonic dystrophy." Journal of Community Hospital Internal Medicine Perspectives **5**(1): 1-5.

Khelifi, M. M., A. Ishmukhametova, P. Khau Van Kien, D. Thorel, D. Mechin, S. Perelman, J. Pouget, M. Claustres and S. Tuffery-Giraud (2011). "Pure intronic

rearrangements leading to aberrant pseudoexon inclusion in dystrophinopathy: a new class of mutations?" Hum Mutat **32**(4): 467-475.

Khurana, T. S., R. A. Prendergast, H. S. Alameddine, F. M. Tome, M. Fardeau, K. Arahata, H. Sugita and L. M. Kunkel (1995). "Absence of extraocular muscle pathology in Duchenne's muscular dystrophy: role for calcium homeostasis in extraocular muscle sparing." J Exp Med **182**(2): 467-475.

Kitano, S., Y. Kino, Y. Yamamoto, M. Takitani, J. Miyoshi, T. Ishida, Y. Saito, K. Arima and J. Satoh (2015). "Bioinformatics Data Mining Approach Suggests Coexpression of AGTPBP1 with an ALS-linked Gene C9orf72." J Cent Nerv Syst Dis **7**: 15-26.

Klein, C. J., D. D. Coovert, D. E. Bulman, P. N. Ray, J. R. Mendell and A. H. Burghes (1992). "Somatic reversion/suppression in Duchenne muscular dystrophy (DMD): evidence supporting a frame-restoring mechanism in rare dystrophin-positive fibers." Am J Hum Genet **50**(5): 950-959.

Klinck, R., A. Fourrier, P. Thibault, J. Toutant, M. Durand, E. Lapointe, M. L. Caillet-Boudin, N. Sergeant, G. Gourdon, G. Meola, D. Furling, J. Puymirat and B. Chabot (2014). "RBFOX1 cooperates with MBNL1 to control splicing in muscle, including events altered in myotonic dystrophy type 1." PLoS One **9**(9): e107324.

Lai, Y., G. D. Thomas, Y. Yue, H. T. Yang, D. Li, C. Long, L. Judge, B. Bostick, J. S. Chamberlain, R. L. Terjung and D. Duan (2009). "Dystrophins carrying spectrin-like repeats 16 and 17 anchor nNOS to the sarcolemma and enhance exercise performance in a mouse model of muscular dystrophy." J Clin Invest **119**(3): 624-635.

Legrand, B., E. Giudice, A. Nicolas, O. Delalande and E. Le Rumeur (2011). "Computational study of the human dystrophin repeats: interaction properties and molecular dynamics." PLoS One **6**(8): e23819.

Lee, D., O. Redfern and C. Orengo (2007). "Predicting protein function from sequence and structure." Nat Rev Mol Cell Biol **8**(12): 995-1005.

Leiden Muscular Dystrophy Pages (2006). "Dystrophin Isoforms." Retrieved 30/09/16, 2016, from www.dmd.nl.

Li, H., D. Chen and J. Zhang (2012). "Analysis of intron sequence features associated with transcriptional regulation in human genes." PLoS One **7**(10): e46784.

Lin, L., W. X. Gu, G. Ozisik, W. S. To, C. J. Owen, J. L. Jameson and J. C. Achermann (2006). "Analysis of DAX1 (NR0B1) and steroidogenic factor-1 (NR5A1) in children and adults with primary adrenal failure: ten years' experience." J Clin Endocrinol Metab **91**(8): 3048-3054.

Lindlof, M., H. Kaariainen, K. E. Davies and A. Delachapelle (1986). "Carrier Detection and Prenatal-Diagnosis in X-Linked Muscular-Dystrophy Using

Restriction-Fragment-Length-Polymorphisms." Journal of Medical Genetics **23**(6): 560-572.

Lu, Q. L., G. E. Morris, S. D. Wilton, T. Ly, O. V. Artem'yeva, P. Strong and T. A. Partridge (2000). "Massive idiosyncratic exon skipping corrects the nonsense mutation in dystrophic mouse muscle and produces functional revertant fibers by clonal expansion." J Cell Biol **148**(5): 985-996.

Lu, X., N. A. Timchenko and L. T. Timchenko (1999). "Cardiac elav-type RNA-binding protein (ETR-3) binds to RNA CUG repeats expanded in myotonic dystrophy." Hum Mol Genet **8**(1): 53-60.

Madden, H. R., S. Fletcher, M. R. Davis and S. D. Wilton (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.

Magri, F., R. Del Bo, M. G. D'Angelo, A. Govoni, S. Ghezzi, S. Gandossini, M. Sciacco, P. Ciscato, A. Bordoni, S. Tedeschi, F. Fortunato, V. Lucchini, M. Cereda, S. Corti, M. Moggio, N. Bresolin and G. P. Comi (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.

Malueka, R. G., Y. Takaoka, M. Yagi, H. Awano, T. Lee, E. K. Dwianingsih, A. Nishida, Y. Takeshima and M. Matsuo (2012). "Categorization of 77 dystrophin exons into 5 groups by a decision tree using indexes of splicing regulatory factors as decision markers." BMC Genet **13**: 23

Mann, C. J., K. Honeyman, A. J. Cheng, T. Ly, F. Lloyd, S. Fletcher, J. E. Morgan, T. A. Partridge and S. D. Wilton (2001). "Antisense-induced exon skipping and synthesis of dystrophin in the mdx mouse." Proc Natl Acad Sci U S A **98**(1): 42-47.

Marchuk, D. A., A. M. Saulino, R. Tavakkol, M. Swaroop, M. R. Wallace, L. B. Andersen, A. L. Mitchell, D. H. Gutmann, M. Boguski and F. S. Collins (1991). "cDNA cloning of the type 1 neurofibromatosis gene: complete sequence of the NF1 gene product." Genomics **11**(4): 931-940.

Matsumura, K., I. Nonaka, F. M. S. Tome, K. Arahata, H. Collin, F. Leturcq, D. Recan, J. C. Kaplan, M. Fardeau and K. P. Campbell (1993). "Mild Deficiency of Dystrophin-Associated Proteins in Becker Muscular-Dystrophy Patients Having in-Frame Deletions in the Rod Domain of Dystrophin." American Journal of Human Genetics **53**(2): 409-416.

Messiaen, L. M., T. Callens, G. Mortier, D. Beysen, I. Vandenbroucke, N. Van Roy, F. Speleman and A. D. Paepe (2000). "Exhaustive mutation analysis of the NF1 gene allows identification of 95% of mutations and reveals a high frequency of unusual splicing defects." Hum Mutat **15**(6): 541-555.

Metzenberg, A. B., G. Wurzer, T. H. Huisman and O. Smithies (1991). "Homology requirements for unequal crossing over in humans." Genetics **128**(1): 143-161.

Mitrpant, C., A. M. Adams, P. L. Meloni, F. Muntoni, S. Fletcher and S. D. Wilton (2009). "Rational design of antisense oligomers to induce dystrophin exon skipping." Mol Ther **17**(8): 1418-1426.

Monaco, A. P., C. J. Bertelson, S. Liechti-Gallati, H. Moser and L. M. Kunkel (1988). "An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus." Genomics **2**(1): 90-95.

Muntoni, F., S. Torelli and A. Ferlini (2003). "Dystrophin and mutations: one gene, several proteins, multiple phenotypes." Lancet Neurology **2**(12): 731-740.

Nakamori, M., T. Kimura, H. Fujimura, M. P. Takahashi and S. Sakoda (2007). "Altered mRNA splicing of dystrophin in type 1 myotonic dystrophy." Muscle Nerve **36**(2): 251-257.

Nakamura, A. (2015). "X-Linked Dilated Cardiomyopathy: A Cardiospecific Phenotype of Dystrophinopathy." Pharmaceuticals (Basel) **8**(2): 303-320.

Nakamura, A., K. Yoshida, K. Fukushima, H. Ueda, N. Urasawa, J. Koyama, Y. Yazaki, M. Yazaki, T. Sakai, S. Haruta, S. Takeda and S. Ikeda (2008). "Follow-up of three patients with a large in-frame deletion of exons 45-55 in the Duchenne muscular dystrophy (DMD) gene." J Clin Neurosci **15**(7): 757-763.

Nardes, F., A. P. Araujo and M. G. Ribeiro (2012). "Mental retardation in Duchenne muscular dystrophy." J Pediatr (Rio J) **88**(1): 6-16.

National Human Genome Research Institute (2016). “The Cost of Sequencing a Human Genome”. Retrieved from: https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/

Nishi, T., P. S. Lee, K. Oka, V. A. Levin, S. Tanase, Y. Morino and H. Saya (1991). "Differential expression of two types of the neurofibromatosis type 1 (NF1) gene transcripts related to neuronal differentiation." Oncogene **6**(9): 1555-1559.

Nouri, N., E. Fazel-Najafabadi, M. Behnam, N. Nouri, O. Aryani, M. Ghasemi, J. Nasiri and M. Sedghi (2014). "Use of in silico tools for classification of novel missense mutations identified in dystrophin gene in developing countries." Gene **535**(2): 250-254.

Oldfors, A., B. O. Eriksson, M. Kyllerman, T. Martinsson and J. Wahlstrom (1994). "Dilated cardiomyopathy and the dystrophin gene: an illustrated review." Br Heart J **72**(4): 344-348.

Pan, Q., O. Shai, L. J. Lee, B. J. Frey and B. J. Blencowe (2008). "Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing." Nat Genet **40**(12): 1413-1415.

Paul, S., W. Dansithong, D. Kim, J. Rossi, N. J. G. Webster, L. Comai and S. Reddy (2006). "Interaction of musleblind, CUG-BP1 and hnRNP H proteins in DM1-associated aberrant IR splicing." Embo Journal **25**(18): 4271-4283.

Passamano, L., A. Taglia, A. Palladino, E. Viggiano, P. D'Ambrosio, M. Scutifero, M. Rosaria Cecio, V. Torre, D. E. L. F, E. Picillo, O. Paciello, G. Piluso, G. Nigro and L. Politano (2012). "Improvement of survival in Duchenne Muscular Dystrophy: retrospective analysis of 835 patients." Acta Myol **31**(2): 121-125.

Perrin, G., M. A. Morris, S. E. Antonarakis, E. Boltshauser and P. Hutter (1996). "Two novel mutations affecting mRNA splicing of the neurofibromatosis type 1 (NF1) gene." Hum Mutat **7**(2): 172-175.

Petrof, B. J., J. B. Shrager, H. H. Stedman, A. M. Kelly and H. L. Sweeney (1993). "Dystrophin protects the sarcolemma from stresses developed during muscle contraction." Proc Natl Acad Sci U S A **90**(8): 3710-3714.

Piva, F., M. Giulietti, A. B. Burini and G. Principato (2012). "SpliceAid 2: a database of human splicing factors expression data and RNA target motifs." Hum Mutat **33**(1): 81-85.

Pramono, Z. A., K. B. Wee, J. L. Wang, Y. J. Chen, Q. B. Xiong, P. S. Lai and W. C. Yee (2012). "A prospective study in the rational design of efficient antisense oligonucleotides for exon skipping in the DMD gene." Hum Gene Ther **23**(7): 781-790.

Prior, T. W., A. C. Papp, P. J. Snyder, A. H. Burghes, C. Bartolo, M. S. Sedra, L. M. Western and J. R. Mendell (1993). "A missense mutation in the dystrophin gene in a Duchenne muscular dystrophy patient." Nat Genet **4**(4): 357-360.

Pros, E., C. Gomez, T. Martin, P. Fabregas, E. Serra and C. Lazaro (2008). "Nature and mRNA effect of 282 different NF1 point mutations: focus on splicing alterations." Hum Mutat **29**(9): E173-193.

Pros, E., J. Fernandez-Rodriguez, B. Canet, L. Benito, A. Sanchez, A. Benavides, F. J. Ramos, M. A. Lopez-Ariztegui, G. Capella, I. Blanco, E. Serra and C. Lazaro (2009). "Antisense therapeutics for neurofibromatosis type 1 caused by deep intronic mutations." Hum Mutat **30**(3): 454-462.

Rafael, J. A., G. A. Cox, K. Corrado, D. Jung, K. P. Campbell and J. S. Chamberlain (1996). "Forced expression of dystrophin deletion constructs reveals structure-function correlations." J Cell Biol **134**(1): 93-102.

Ranum, L. P. and T. A. Cooper (2006). "RNA-mediated neuromuscular disorders." Annu Rev Neurosci **29**: 259-277.

Raponi, M., M. Upadhyaya and D. Baralle (2006). "Functional splicing assay shows a pathogenic intronic mutation in neurofibromatosis type 1 (NF1) due to intronic sequence exonization." Hum Mutat **27**(3): 294-295.

Rau, F., J. Laine, L. Ramanoudjame, A. Ferry, L. Arandel, O. Delalande, A. Jollet, F. Dingli, K. Y. Lee, C. Peccate, S. Lorain, E. Kabashi, T. Athanasopoulos, T. Koo, D. Loew, M. S. Swanson, E. Le Rumeur, G. Dickson, V. Allamand, J. Marie and D. Furling (2015). "Abnormal splicing switch of DMD's penultimate exon compromises muscle fibre maintenance in myotonic dystrophy." Nat Commun **6**: 7205.

Rearick, D., A. Prakash, A. McSweeny, S. S. Shepard, L. Fedorova and A. Fedorov (2011). "Critical association of ncRNA with introns." Nucleic Acids Res **39**(6): 2357-2366.

Roberts, R. G., D. R. Bentley and M. Bobrow (1993). "Infidelity in the structure of ectopic transcripts: a novel exon in lymphocyte dystrophin transcripts." Hum Mutat **2**(4): 293-299.

Ruszczak, C., A. Mirza and N. Menhart (2009). "Differential stabilities of alternative exon-skipped rod motifs of dystrophin." Biochim Biophys Acta **1794**(6): 921-928.

Rybakova, I. N., K. J. Amann and J. M. Ervasti (1996). "A new model for the interaction of dystrophin with F-actin." J Cell Biol **135**(3): 661-672.

Rybakova, I. N., J. R. Patel and J. M. Ervasti (2000). "The dystrophin complex forms a mechanically strong link between the sarcolemma and costameric actin." Journal of Cell Biology **150**(5): 1209-1214.

Sakharkar, M. K., V. T. Chow and P. Kangueane (2004). "Distributions of exons and introns in the human genome." In Silico Biol **4**(4): 387-393.

Schara, U. and B. G. Schoser (2006). "Myotonic dystrophies type 1 and 2: a summary on current aspects." Semin Pediatr Neurol **13**(2): 71-79.

Senapathy, P., M. B. Shapiro and N. L. Harris (1990). "Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project." Methods Enzymol **183**: 252-278.

Sfeir, A. and L. S. Symington (2015). "Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway?" Trends Biochem Sci **40**(11): 701-714.

Shepard, P., E. M. Lam, E. K. St Louis and J. Dominik (2012). "Sleep disturbances in myotonic dystrophy type 2." Eur Neurol **68**(6): 377-380.

Sherratt, T. G., T. Vulliamy, V. Dubowitz, C. A. Sewry and P. N. Strong (1993). "Exon skipping and translation in patients with frameshift deletions in the dystrophin gene." Am J Hum Genet **53**(5): 1007-1015.

Shin, C., F. E. Kleiman and J. L. Manley (2005). "Multiple properties of the splicing repressor SRp38 distinguish it from typical SR proteins." Mol Cell Biol **25**(18): 8334-8343.

Shin, J., M. M. Tajrishi, Y. Ogura and A. Kumar (2013). "Wasting mechanisms in muscular dystrophy." Int J Biochem Cell Biol **45**(10): 2266-2279.

Sironi, M., G. Menozzi, L. Riva, R. Cagliani, G. P. Comi, N. Bresolin, R. Giorda and U. Pozzoli (2004). "Silencer elements as possible inhibitors of pseudoexon splicing." Nucleic Acids Res **32**(5): 1783-1791.

Siva, K., G. Covello and M. A. Denti (2014). "Exon-skipping antisense oligonucleotides to correct missplicing in neurogenetic diseases." Nucleic Acid Ther **24**(1): 69-86.

Skandalis, A. (2016). "Estimation of the minimum mRNA splicing error rate in vertebrates." Mutat Res **784-785**: 34-38.

Smith, P. J., C. Zhang, J. Wang, S. L. Chew, M. Q. Zhang and A. R. Krainer (2006). "An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers." Hum Mol Genet **15**(16): 2490-2508.

Stankiewicz, P. and J. R. Lupski (2002). "Genome architecture, rearrangements and genomic disorders." Trends in Genetics **18**(2): 74-82.

Straub, V. and K. P. Campbell (1997). "Muscular dystrophies and the dystrophin-glycoprotein complex." Current Opinion in Neurology **10**: 168-175.

Sun, H. and L. A. Chasin (2000). "Multiple splicing defects in an intronic false exon." Mol Cell Biol **20**(17): 6414-6425.

Suminaga, R., Y. Takeshima, K. Adachi, M. Yagi, H. Nakamura and M. Matsuo (2002). "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 **47**(4): 196-201.

Takeshima, Y., M. Yagi, Y. Okizuka, H. Awano, Z. Zhang, Y. Yamauchi, H. Nishio and M. Matsuo (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.

Tieleman, A. A., J. van Viet, J. B. M. J. Jansen, A. J. van der Kooi, G. F. Borm and B. G. M. van Engelen (2008). "Gastrointestinal involvement is frequent in Myotonic Dystrophy type 2." Neuromuscular Disorders **18**(8): 646-649.

Timchenko, N. A., R. Patel, P. Iakova, Z. J. Cai, L. Quan and L. T. Timchenko (2004). "Overexpression of CUG triplet repeat-binding protein, CUGBP1, in mice inhibits myogenesis." J Biol Chem **279**(13): 13129-13139.

Trabelsi, M., C. Beugnet, N. Deburgrave, V. Commere, L. Orhant, F. Leturcq and J. Chelly (2014). "When a mid-intronic variation of DMD gene creates an ESE site." Neuromuscular Disorders **24**(12): 1111-1117.

Tran, V. K., Z. Zhang, M. Yagi, A. Nishiyama, Y. Habara, Y. Takeshima and M. Matsuo (2005). "A novel cryptic exon identified in the 3' region of intron 2 of the human dystrophin gene." J Hum Genet **50**(8): 425-433.

Tuffery-Giraud, S., C. Saquet, S. Chambert and M. Claustres (2003). "Pseudoexon activation in the DMD gene as a novel mechanism for Becker muscular dystrophy." Hum Mutat **21**(6): 608-614.

van den Bergen, J. C., B. H. Wokke, A. A. Janson, S. G. van Duinen, M. A. Hulsker, H. B. Ginjaar, J. C. van Deutekom, A. Aartsma-Rus, H. E. Kan and J. J. Verschuuren (2014). "Dystrophin levels and clinical severity in Becker muscular dystrophy patients." J Neurol Neurosurg Psychiatry **85**(7): 747-753.

Verdin, H., B. D'Haene, D. Beysen, Y. Novikova, B. Menten, T. Sante, P. Lapunzina, J. Nevado, C. M. Carvalho, J. R. Lupski and E. De Baere (2013). "Microhomology-mediated mechanisms underlie non-recurrent disease-causing microdeletions of the FOXL2 gene or its regulatory domain." PLoS Genet **9**(3): e1003358.

Vihinen, M. (2014). "Majority Vote and Other Problems when using Computational Tools." Human Mutation **35**(8): 912-914.

Vissers, L. E., S. S. Bhatt, I. M. Janssen, Z. Xia, S. R. Lalani, R. Pfundt, K. Derwinska, B. B. de Vries, C. Gilissen, A. Hoischen, M. Nesteruk, B. Wisniowiecka-Kowalnik, M. Smyk, H. G. Brunner, S. W. Cheung, A. G. van Kessel, J. A. Veltman and P. Stankiewicz (2009). "Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genomic architecture." Hum Mol Genet **18**(19): 3579-3593.

Wang, E. T., A. J. Ward, J. M. Cherone, J. Giudice, T. T. Wang, D. J. Treacy, N. J. Lambert, P. Freese, T. Saxena, T. A. Cooper and C. B. Burge (2015). "Antagonistic regulation of mRNA expression and splicing by CELF and MBNL proteins." Genome Res **25**(6): 858-871.

Wang, Z., M. Gerstein and M. Snyder (2009). "RNA-Seq: a revolutionary tool for transcriptomics." Nat Rev Genet **10**(1): 57-63.

Watermann, D. O., Y. Tang, A. Z. Hausen, M. Jäger, S. Stamm and E. Stickeler (2006). "Splicing factor Tra2-β1 is specifically induced in breast cancer and regulates alternative splicing of the CD44 gene." Cancer Research **66**(9): 4774-4780.

Weise, A., K. Mrasek, E. Klein, M. Mulatinho, J. C. Llerena, Jr., D. Hardekopf, S. Pekova, S. Bhatt, N. Kosyakova and T. Liehr (2012). "Microdeletion and microduplication syndromes." J Histochem Cytochem **60**(5): 346-358.

Wilton, S. D. and S. Fletcher (2010). "Splice modification to restore functional dystrophin synthesis in Duchenne muscular dystrophy." Curr Pharm Des **16**(8): 988-1001.

Wilton, S. (2013). "An Update on Dmd Exon Skipping Trials: Making More Sense with Splice Switching Antisense Oligonucleotides." Clinical and Experimental Pharmacology and Physiology **40**: 38-38.

Wilton, S. D., A. M. Fall, P. L. Harding, G. McClorey, C. Coleman and S. Fletcher (2007). "Antisense oligonucleotide-induced exon skipping across the human dystrophin gene transcript." Mol Ther **15**(7): 1288-1296.

Wilton, S. D., L. Lim, D. Dye and N. Laing (1997). "Bandstab: a PCR-based alternative to cloning PCR products." Biotechniques **22**(4): 642-645.

Wimmer, K., M. Eckart, H. Rehder and C. Fonatsch (2000). "Illegitimate splicing of the NF1 gene in healthy individuals mimics mutation-induced splicing alterations in NF1 patients." Hum Genet **106**(3): 311-313.

Wiszniewska, J., W. Bi, C. Shaw, P. Stankiewicz, S. H. Kang, A. N. Pursley, S. Lalani, P. Hixson, T. Gambin, C. H. Tsai, H. G. Bock, M. Descartes, F. J. Probst, F. Scaglia, A. L. Beaudet, J. R. Lupski, C. Eng, S. W. Cheung, C. Bacino and A. Patel (2014). "Combined array CGH plus SNP genome analyses in a single assay for optimized clinical testing." Eur J Hum Genet **22**(1): 79-87.

Wu, B., P. Lu, C. Cloer, M. Shaban, S. Grewal, S. Milazi, S. N. Shah, H. M. Moulton and Q. L. Lu (2012). "Long-term rescue of dystrophin expression and improvement in muscle pathology and function in dystrophic mdx mice by peptide-conjugated morpholino." Am J Pathol **181**(2): 392-400.

Yagi, M., Y. Takeshima, H. Wada, H. Nakamura and M. Matsuo (2003). "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." Hum Genet **112**(2): 164-170.

Yan, H. B., Z. Z. Lou, L. Li, P. J. Brindley, Y. D. Zheng, X. N. Luo, J. L. Hou, A. J. Guo, W. Z. Jia and X. P. Cai (2014). "Genome-wide analysis of regulatory proteases sequences identified through bioinformatics data mining in Taenia solium." Bmc Genomics **15**.

Yang, Y. H., M. A. Markus, A. H. Mangs, O. Raitskin, R. Sperling and B. J. Morris (2013). "ZRANB2 localizes to supraspliceosomes and influences the alternative splicing of multiple genes in the transcriptome." Mol Biol Rep **40**(9): 5381-5395.

Yang, Y., L. Zhan, W. Zhang, F. Sun, W. Wang, N. Tian, J. Bi, H. Wang, D. Shi, Y. Jiang, Y. Zhang and Y. Jin (2011). "RNA secondary structure in mutually exclusive splicing." Nat Struct Mol Biol **18**(2): 159-168.

Yokota, T., Q. L. Lu, J. E. Morgan, K. E. Davies, R. Fisher, S. Takeda and T. A. Partridge (2006). "Expansion of revertant fibers in dystrophic mdx muscles reflects activity of muscle precursor cells and serves as an index of muscle regeneration." J Cell Sci **119**(Pt 13): 2679-2687.

Zhang, Z., Y. Habara, A. Nishiyama, Y. Oyazato, M. Yagi, Y. Takeshima and M. Matsuo (2007). "Identification of seven novel cryptic exons embedded in the dystrophin gene and characterization of 14 cryptic dystrophin exons." J Hum Genet **52**(7): 607-617.