5-2022

Development of a Long-Read Sequencing Protocol to Assess the Precision and Efficacy of Gene Editing for Duchenne Muscular Dystrophy

Landon Andrew Burcham
University of Arkansas, Fayetteville

Development of a Long-Read Sequencing Protocol to Assess the Precision and Efficacy of Gene Editing for Duchenne Muscular Dystrophy

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Biomedical Engineering

by

Landon Burcham
Harding University
Bachelor of Science in Biomedical Engineering, 2019

May 2022
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

_____
Christopher Nelson, Ph.D.
Thesis Director

_____
Jeffery Wolchok, Ph.D.
Committee Member

_____
Jin-Woo Kim, Ph.D.
Committee Member

Abstract

This work establishes a method for assessing on-target precision due to CRISPR-Cas9 gene editing, especially within the context of exon skipping therapy for Duchenne Muscular Dystrophy. The proposed method utilizes an Oxford nanopore long-read sequencing approach to sequence amplified regions of DNA that have been edited using CRISPR-Cas9. NIH3T3 and C2C12 cell lines were treated with a dual-guide CRISPR-Cas9 system, that targets and deletes exon 23 from the DMD gene in mouse samples. Deletion PCR revealed deletion of exon 23 in both DNA and cDNA samples. Additionally, sequencing using Oxford Nanopore revealed targeted exon 23 deletion as the most prevalent event. There were additional editing events, including large deletions, insertions, pseudoexons, and a duplication. The next step is to sequence in vivo mouse samples that were treated with the same dual-guide system targeting exon 23. This will reveal additional unintended editing events such as vector integration and will provide further insight into the efficacy of CRISPR-based therapies. Future work will focus on optimizing an amplification-free method for target enrichment and long-read sequencing.

Table of Contents

Chapter 1: Literature Review

1.0 Introduction

The discovery of CRISPR-Cas9, as a tool for gene editing, has resulted in potentially transformative research that has led to a variety of novel treatment approaches for some of the most severe genetic diseases (Xiao-Jie et al., 2015). The basic premise of therapeutic gene editing with CRISPR-Cas9 is that it can target and alter a variety of genomic regions, enabling the removal or correction of mutated regions of DNA, thus restoring partial or fully functional genes of relevance in a variety of genetic diseases (Cong et al., 2013; Hsu et al., 2014; Jinek et al., 2013). However, safety and efficacy concerns remain a barrier to clinical translation. One of the main concerns regarding the safety and efficacy of CRISPR-based gene editing is the unintended on-target and off-target alterations (You et al., 2019). In order to evaluate the long-term safety and efficacy of gene editing, there needs to be a strong research focus on identifying the shortcomings of CRISPR-based gene therapies. This includes quantifying unintended genome modifications using unbiased methods and understanding the consequences of genomic changes on transcript isoform diversity. The rapid advancement of next-generation sequencing technologies has enabled the detection of unintended genomic modifications caused by CRISPR-based therapeutic approaches.

These are basic questions that, if answered, will guide the development of gene editing therapies in numerous diseases. It is important to detect these unintentional changes before clinical translation. The proposed work provides a foundation for future long-read sequencing applications and is guided by two basic aims:

1. To transfect mouse cells with a dual-guide CRISPR-Cas9 system and analyze on-target cleavage efficiency using gel-based methods

2. To develop a long-read sequencing workflow to characterize unintentional modifications induced by the CRISPR-based therapeutic for Duchenne Muscular Dystrophy (DMD)

1.1 Duchenne Muscular Dystrophy

Duchenne Muscular Dystrophy (DMD) is an X-linked recessive monogenic disorder that is caused by mutations in the dystrophin gene which encodes a membrane-associated protein that is essential for muscular function (Nelson, 2019). Dystrophin plays an integral role in maintaining sarcolemmal integrity and stability through its interaction with intracellular actin and extracellular laminin (Zhang et al., 2020). Despite being a genetic disorder that affects only one gene, the pathological mechanisms are quite complex. Dystrophin is expressed largely in skeletal muscle, as well as in cardiac, vascular, intestinal smooth muscle, and nervous system tissue (Nakamura, 2019). It affects 1 in 5000 males and the life expectancy is 26 years of age (Nelson, 2019). The principal symptom of DMD is severe muscle weakness, which ultimately leads to cardiomyopathy and compromised respiratory function (Chamberlain & Chamberlain, 2017).

The DMD gene is 2.2 Mb in size with 79 exons, making it the largest known human gene. It has numerous isoforms that are expressed in muscle and non-muscle tissues and displays the highest known spontaneous mutation frequency, likely due to its large size (Chamberlain & Chamberlain, 2017). Deletions account for 60-70% of all mutations resulting in DMD, point mutations are responsible for 26% of mutations, and duplications are responsible for 10-15% of all mutations (Gao & McNally, 2015). Most Duchenne Muscular Dystrophy cases are the result of a disrupted mRNA reading frame or a premature stop codon that terminates translation (Mann et al., 2001). In the less severe Becker Muscular Dystrophy (BMD), the mRNA reading frame is usually maintained, resulting in a semi-functional dystrophin product (Mann et al., 2001).

The milder Becker Muscular Dystrophy usually maintains the reading frame, yielding partial dystrophin expression.

1.2 Therapeutic Approaches for Duchenne Muscular Dystrophy

Various therapeutic strategies have been explored to restore dystrophin function, including myoblast and stem cell transfer, upregulation of utrophin, and exosome injection (Olson, 2021). Corticosteroid-based therapies have been used to reduce inflammation and aid in sarcolemmal repair (Mackenzie et al., 2021). Studies have also shown the role of intermittent dosing in reducing muscle fiber atrophy. The first FDA approved drug for the treatment of DMD is a morpholino-modified oligonucleotide called eteplirsen (Lim et al., 2017). It promotes exon skipping by masking the splice acceptor sequence in exon 51. However, studies have shown inadequate expression of dystrophin, after one year of dosing, to confer clinical benefits and thus other oligonucleotides are being explored (Olson, 2021). Antisense oligonucleotide (AON) therapy has also been used to target exon 23, resulting in a modified dystrophin protein and partial recovery of dystrophin function (Mann et al., 2001) However, antisense oligonucleotide therapy has several disadvantages in that these therapies must be readministered throughout the patient's life and may lead to a dystrophin protein with an abnormal spectrin-like repeats (SLR) (Tremblay et al., 2016).

Adeno-associated virus (AAV) vectors that demonstrate a lack of pathogenicity and systemic expression in muscle cells have been utilized in systemic gene therapies (Yue et al., 2008). However, the limited packaging size of AAV vectors remains an obstacle, only allowing gene sizes smaller than 4.5 kilobases (Wang et al., 2000). This limitation excludes large genes such as dystrophin, with a 14 kilobase complementary DNA. Thus, efforts have been made to construct truncated dystrophins that restore a level of muscular function in canine and mouse models (Shin

et al., 2013). This approach stems from an understanding of the milder Becker Muscular Dystrophy (BMD), which results from a truncated dystrophin gene. Prior studies have shown that a 6-8 kb mini dystrophin has been effective but still exceeds the packaging limits of AAV (Duan, 2018). Microdystrophin administration has been shown to effectively reduce muscle disease in both mdx mouse models and dystrophic dog models (Shin et al., 2013). CRISPR-based methods have been offered, in recent years, as an alternative to AON therapy and is explored in the next section. There are ongoing clinical trials by Sarepta, Pfizer, and Solid Biosciences demonstrating the clinical utility and safety of these microdystrophin-based approaches (Mendell et al., 2020; Schneider et al., 2017).

1.2.1 CRISPR Cas9 System

The earliest discovery of CRISPR dates to the late 1980s with the identification of repeating sequences found in Escherichia coli (Ishino et al., 1987). These repeats were found in a range of other bacteria, suggesting the importance and evolutionary conservation of these repeats (Ishino et al., 2018). Crucial to the understanding of CRISPR was a recognition of the relation to cas genes. A paper published in 2005 by Mojica et al. revealed sequence similarity between the spacer regions of CRISPR and the sequences of viruses and bacteriophages, identifying an immune function in eukaryotes (Mojica et al., 2005). Furthermore, CRISPR and its association with cas genes was identified to play a role in protecting prokaryotes from invading viruses and bacteria (Ishino et al., 2018).

Several pivotal papers released in 2012 describing CRISPR Cas9 as a programmable nuclease for gene editing has revolutionized treatment of fatal genetic disease as well as applications in diagnostics and epigenetic regulation (Hsu et al., 2014; Jinek et al., 2013). In the context of

Duchenne Muscular Dystrophy, CRISPR-Cas9 based therapies have corrected normal phenotype in cells and mice and restored a level of dystrophin expression in human cells.

CRISPR-Cas9 is a revolutionary molecular tool that enables the precise correction of genetic mutations that are responsible for DMD, with the possibility of permanently restoring muscular function and other pathological consequences of the disease (Olson, 2021).

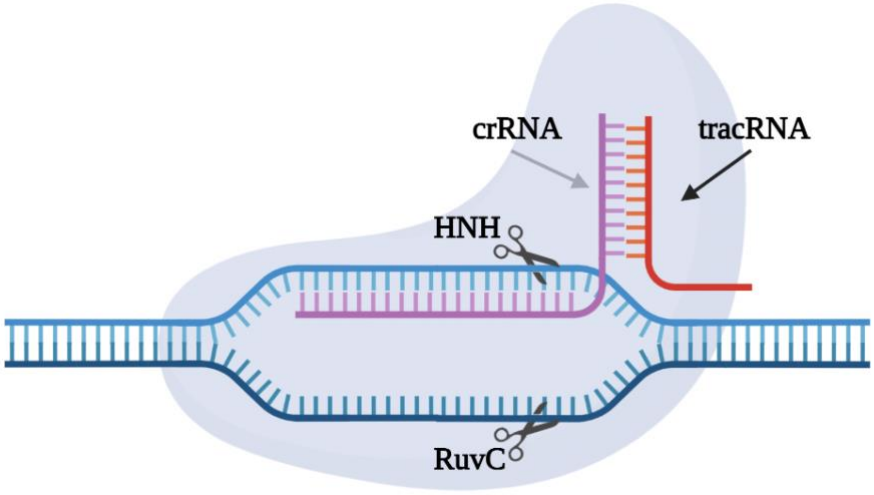

Figure 1.1: CRISPR-Cas9 (Created using Biorender.com)

A single guide RNA (sgRNA) is designed to target a specific region of the DNA, where the Cas9 protein induces a double-stranded break. The sgRNA is a fusion of a trans-activating RNA (tracrRNA), which is a non-coding RNA that facilitates crRNA-guided cleavage by Cas9 and serves as a binding scaffold for Cas nuclease (Jinek et al., 2012). The CRISPR RNA (crRNA) contains a 20-base pair sequence that is complementary to the DNA target sequence. The correct functioning of the Cas9 system is also dependent upon the presence of a protospacer adjacent motif (PAM) that is an appropriate distance from the sgRNA target site. The PAM site differs based on the bacterial species (eg. 5'NGG3' for streptococcus *pyrogenes* Cas9) of the Cas9 gene and is found directly downstream of the target sequence. After the target sequence has been identified and matches the sgRNA sequence, the HNH nuclease domain of the Cas nuclease

cleaves the target strand while the RuvC-like nuclease domain of the Cas helical lobe cleaves the non-target strand (Jinek et al., 2012).

CRISPR-Cas9 gene editing occurs through one of two mechanisms: Homology-directed repair (HDR) or nonhomologous end joining (NHEJ). Homology-directed repair requires an exogenous DNA template and generates precise edits at a target locus. This mechanism is limited to proliferating cells and cannot be applied to DMD correction because of the post-mitotic nature of muscle cells. One of the early studies, demonstrating the in vivo potential of CRISPR as a therapeutic for DMD, aimed to correct nonsense mutation at exon 23 through NHEJ and HDR pathways using a single gRNA. The components were injected into mdx mice at the 1-cell stage and implanted into pseudopregnant mice. Immunostaining revealed near wild-type expression levels in pups with 83% NHEJ or 41% HDR correction at 7-9 weeks of age (Lim et al., 2018). Furthermore, 47-60% of muscle fibers showed dystophin-postive staining with only a 17% HDR correction.

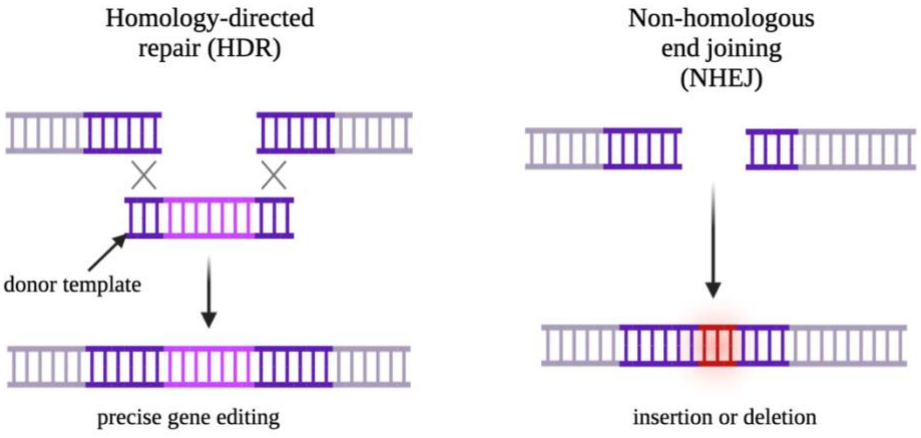

Figure 1.2: CRISPR-Cas9 repair mechanisms (Created using Biorender.com)

This method has proven challenging, due to both the low rate of HDR in post-mitotic cells and the ethical issues surrounding germline editing strategies. Thus, most efforts have been focused on applying the NHEJ mechanism to somatic cells.

IN NHEJ repair, Cas9-induced DNA cutting results in insertions or deletions (INDEL) in the absence of an exogenous DNA template (Olson, 2021). This method is not as precise as HDR and is more prone to error. This editing strategy can be applied to induce either a single-cut or double-cut using either a single sgRNA or pair of sgRNAs. These induced DSBs restore the open-reading frame by exon skipping (Nakamura, 2019). Exon skipping therapy hinges upon the basic understanding that there are exons that, if skipped, do not disrupt the reading frame, and the translated product will result in partial recovery of dystrophin protein. Unlike other monogenic disorders, full restoration of dystrophin function is not necessary for a therapy to be deemed effective. Research has shown that as little as 15% of normal levels of dystrophin expression is necessary to confer clinical benefits (Chemello et al., 2021b). In the case of exon skipping therapy for exon 23 in DMD, the slightly shortened dystrophin has the potential to minimize the severity of DMD (Mann et al., 2001). Studies have shown that local and systemic delivery of a CRISPR-Cas9 complex targeting exon 23 in mdx mice has resulted in expression of the modified dystrophin gene, improvements in muscle biochemistry, and increased muscle force (Amoasii et al., 2018; Nelson et al., 2016; Tabebordbar et al., 2016).

1.3 Safety Issues of CRISPR-Cas9 therapy

Studies into CRISPR-Cas9 therapy have demonstrated the amelioration of disease phenotypes in both in vivo and in vitro models. However, there are a variety of concerns to consider as these therapies move toward clinical applications. Namely, there have been concerns about the long-term efficacy of these therapies as most studies have been conducted over relatively short periods of time.

One of the central concerns surrounding CRISPR-based gene therapy is the unintended off- and on-target edits. Some of these unintended effects include large deletions, inversions, and viral

vector integration where the viral genome integrates into the double strand break site (Nelson et al., 2019). There are also concerns that there are aberrant splicing events at sites upstream and downstream of the cut site. Several off-target effects that have been observed include the incorporation of exogenous DNA, large (kilobase) chromosomal deletions, and chromosomal translocations (Hendel et al., 2015). Aberrant splicing events were also observed by targeted RNA sequencing. Multi-exon skipping has also been observed by other groups because of gene editing, which likely results from disruption of exon splicing enhancers. It is known that CRISPR gene editing induces alternative splicing of mRNAs, which could lead to multiple RNA isoforms that result from a single genome editing approach. The RNA isoforms are largely unpredictable and further exploration of isoform diversity will help explain the unexpected editing outcomes that occur due to CRISPR genome editing.

1.4 Next-generation Sequencing to Characterize CRISPR Editing Efficiency

Next-generation sequencing strategies have previously been employed to characterize gene editing. Conventional next-generation sequencing methods, shown in Table 1.3, have been utilized for its single base accuracy and thus its ability to resolve single-nucleotide variations or indels. However, these sequencing-based methods have intrinsic biases that result in undetected gene-editing outcomes or poor quantification of these outcomes. Resequencing of short amplicons using next-generation sequencing has often resulted in editing events going undetected. Careful experimental design has documented previously undetected genome edits in therapeutic development for LCA and DMD (Maeder et al., 2019; Nelson et al., 2019).

Method	Advantages/Utility	Insensitive to...	References
Whole genome	SNV, small indels, frequent structural changes	Low frequency changes	Petersen et al, 2017. Belkadi et al, 2015.
Whole exome	Less costly than WGS	Large structural variations	Petersen et al, 2017. Warr et al, 2015.
RNA-seq	Differential expression	Low expression splicing isoforms	Corchete et al, 2020.
Probe enrichment	Small structural changes, indels, SNPs	Large structural variations	Gulilat et al, 2019.
Amplicon resequencing	Small indels	Structural changes, prone to PCR bias	Giannoukous et al, 2018. Nelson et al, 2019.
Nanopore sequencing	Cost-effective, efficient, large SVs	SNPs, small indels	Hoijer et al, 2020. Jia et al, 2014, Gilpatrick et al, 2020. Stevens et al, 2019.

Table 1.3: Next-generation sequencing methods

As discussed previously, unexpected editing outcomes have been observed with CRISPR-based gene editing and, as the potential for gene therapy to be administered into human patients increases, it is becoming increasingly important to understand any unintentional consequences. Gel-based assays, such as T7 endonuclease, have been used to assess the nuclease cleavage efficiency but only has a detection limit of 1-2% modified nuclease alleles and does not give insight into the mutation that is introduced by the nuclease (Hendel et al., 2015). PCR-based methods have also been used to sequence an amplicon using either Sanger or next-generation sequencing, but amplification-based methods often underestimate the activity of the engineered nuclease. Nucleases can induce large deletions that stretch beyond the boundaries of the PCR amplicon and will thus not be detected. Thus, long-read DNA sequencing, ranging from 10 kilobases to >1 Mb in length, has been proposed as a method for better understanding how CRISPR induces off-target edits and enables the detection of large structural variations (Logsdon et al., 2020). Long-read sequencing differs from short-read Illumina sequencing in that it can achieve much longer read lengths. Illumina achieves a read length of <300 base pairs, which is

insufficient for detecting larger structural variations.(Logsdon et al., 2020) There are several long-read sequencing technologies: Pacific Biosciences (PacBio) and Oxford Nanopore (ONT). Each of these technologies has underlying differences in terms of sequencing chemistry used and detection approaches (Logsdon et al., 2020). PacBio, one of the two most common long-read sequencing instruments, utilizes a sequencing by synthesis method called single-molecule real time (SMRT) sequencing. This method detects fluorescently tagged nucleotides as they are synthesized along a DNA template molecule and is limited by the longevity of the polymerase (Amarasinghe et al., 2020). This long-read sequencing method can achieve a read length in the range of tens of kilobases, vastly surpassing the 300 kb maximum achieved by Illumina sequencing. The raw base-called error rate has been reduced to <1% with improvements in the technology over time (Amarasinghe et al., 2020). The other most common long-read sequencing technology comes from Oxford Nanopore Technologies. Unlike other long-read sequencing technologies, ONT sequences native strands of DNA by recording changes in electrical current, that correspond to a particular nucleotide base, as it passes through a protein nanopore (Weirather et al., 2017). Of the two most popular sequencing technologies, Oxford Nanopore boasts the longest read lengths, from 500 bp to 2.3 Mb (Amarasinghe et al., 2020).

1.5 Targeted Sequencing Approaches

Targeted sequencing approaches provide an enormous benefit in terms of cost, time and computational resources required for sequencing. Nucleic acid enrichment can be used to selectively target and amplify a specific region of DNA for sequencing. Enrichment is valuable for detecting DNA present at low levels in samples and for processing samples containing a high level of background sequence, as is often the case with diagnostic applications (Schultzhaus et al., 2021). Targeted amplification of large targets using long-range PCR is a fast and cost-

effective method for enriching specified regions of DNA for sequencing. Long-range PCR kits, such as NEB's Long Amp kit, has a modified polymerase that allows for amplicons as high as 30 kilobases (Jia et al., 2014). However, PCR-based methods introduce amplification bias and important data, such as methylation profiles, is lost when the DNA is amplified. Some also argue that PCR is not a true enrichment method because it produces copies rather than native DNA and will likely favor high abundance alleles in a sample (Schultzhaus et al., 2021). Additionally, the size of the target region that can be sequenced is limited with PCR-based methods. With the more widespread sequencing of native DNA sequences, the need for an amplification-free enrichment method has arisen. Thus, amplification-free methods such as hybridization capture and CRISPR approaches are being utilized as an alternative to PCR-based approaches. Hybridization-based capture methods, in the context of next-generation sequencing, have improved upon PCR-based methods by enabling the retrieval of large genomic fragments at a high sequencing coverage (Albert et al., 2007; Gnirke et al., 2009). Hybridization-based methods, while they have improved accuracy and are more cost-effective, still heavily rely on fragmentation of DNA, limiting the ability to sequence large fragments and the ability to classify large structural variants (Dapprich et al., 2016). A Nextera-transposon-based method has been utilized to overcome PCR bias and sensitively detect small rearrangements as well as larger structural variations (Giannoukos et al., 2018). The UDiTas method incorporates a TN5 transposon that simultaneously fragments the DNA and adds Illumina sequencing adapters. The Tn5 transposon contains an Illumina forward adapter, sample barcode, and unique molecule identifier (UMI) which serves to specifically reduce biases associated with several rounds of PCR (Giannoukos et al., 2018).

Chapter 2: Development of and Delivery Dual-Guide CRISPR-Cas9 System

2.1 Introduction

The first component of this project is to prepare plasmid containing the correct guide RNAs and to transfect two different cell lines with the dual guide CRISPR-Cas9 system. The exon 23 deletion technique, developed by Nelson et. al was, was used as an initial proof-of-concept for developing a sequencing workflow to analyze potential unintended alterations (Nelson et al., 2016). The rationale behind this therapeutic approach is to design two guide RNAs, one that targets intron 22 and another that targets intron 24, effectively removing exon 23 and allowing for partial recovery of dystrophin expression.

Lipofection was used to deliver the dual-guide CRISPR-Cas9 system into NIH3T3 cell line. Chemical-based lipofection is well-characterized in gene editing applications and is easy to implement. Commercially available nanoparticle-based transfection kits, such as Lipofectamine, enhance DNA and RNA delivery into cells in vitro (Givens et al., 2018). The dual-guide CRISPR-Cas9 system was also electroporated into C2C12 cells, which are dystrophin expressing when differentiated into myotubes. This will allow for downstream RNA sequencing analysis and characterization of unexpected RNA isoforms. Plasmid transfection, as opposed to other mRNA or ribonucleoprotein (RNP), was used as a starting point because of its stability and ease of preparation for early experiments (Fajrial et al., 2020).

2.2 Materials and Methods

2.2.1 Guide RNA (gRNA) preparation

The gRNA oligos, with sequences in Appendix A and previously designed by Nelson et al for exon 23 deletion, were ordered from Integrated DNA Technologies (IDT). The process of cloning a sgRNA vector is outlined as follows:

1. Digestion of vector with Bbsi restriction enzyme
2. Anneal oligos
3. Ligation annealed oligos into restriction site of digested vector
4. Pick transformed colony and confirm via sequencing

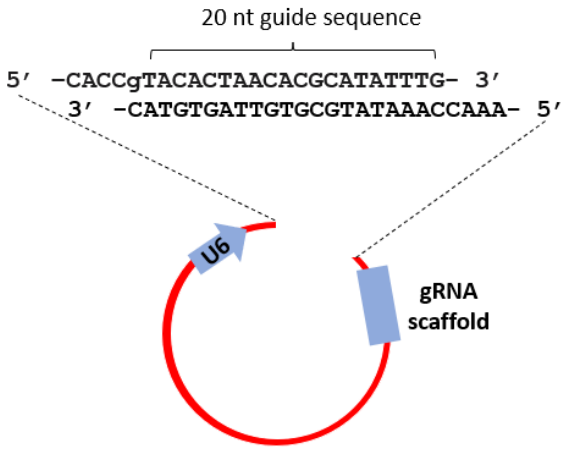

Figure 2.1: Ligation of 20-nt guide sequence into Bbsi restriction site

They were first cloned into a SaCas9 plasmid encoding site (Figure 2.1). Plasmid digestion was performed using the Bbsi restriction enzyme, in combination with 10X buffer, CIP, molecular grade water, and incubated at 37 degrees Celsius for one hour. The digested product was run on a 1% TAE at 100V for 30 minutes. Using a transilluminator for visualization, the gel band was cut out and extracted using the Qiagen Gel Extraction kit.

The primers were then phosphorylated by combining the forward and reverse primers, 2 µL of T4 DNA ligase buffer, 1µL of PNK, and 0.5 µL of ATP. The mixture was put on a PCR block and run under the following conditions:

(1) 37°C for 30-60 min

(2) 65 °C for 20 min

(3) 95 °C for 2 min

(4) Cool to 20 ˚C over 25 min

(5) Hold at 4C

The annealed primers were then ligated into the digested vector using a large excess of plasmid DNA to vector insert. The following ligation mixture was prepared: 50 ng of plasmid DNA, 1 µL insert 1 µL of T4 DNA ligase, 2 µL 10X T4 DNA, water to 20 µL volume. The mixture was held at 16 ˚C on a PCR cycler overnight.

The gRNA vector was then transformed into STBL3 cells and plated onto agar supplemented with Kanamycin and incubated overnight at 37 ˚C. The following day, one of the colonies was inoculated into 5 mL of LB broth supplemented with 1:1000 kanamycin. On the third day, the culture was centrifuged at 4000xg for 10 minutes and plasmid was extracted using the QIAGEN miniprep kit. The plasmid was sequence confirmed and a glycerol stock stored in -80 ˚C for future use.

2.2.2 Delivery of dual-guide CRISPR-Cas9 system in vitro

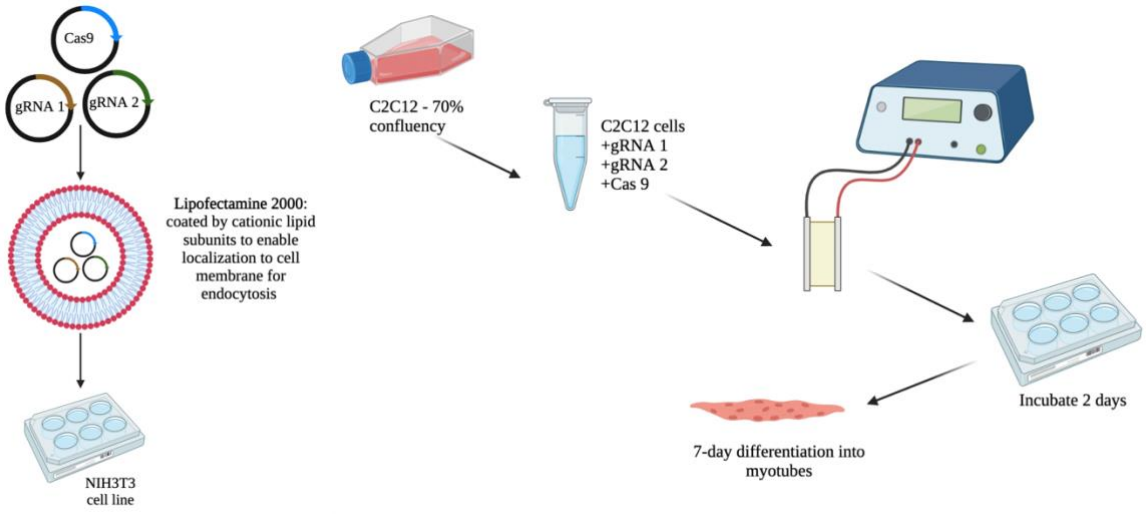

Figure 2.2: Lipid nanoparticle and electroporation-based transfection of gRNA and Cas9 plasmids (Created using Biorender.com)

The NIH3T3 cells were maintained in culture medium supplemented with 10% Fetal Bovine Serum (FBS) and 1% Pen-Strep. The cells were seeded onto a 12-well plate the day before transfection and grown to 70% confluency. For each transfection sample, DNA was diluted in 50 µL of Opti-MEM medium. Cas9:gRNA was mixed in a 1:1 mass ratio. The Lipofectamine 2000 was gently mixed before use and diluted in 50 µL of Opti-MEM Medium. To optimize plasmid transfection, several lipofectamine(µL):plasmid (ng) concentrations were used: 1:2, 1:3, 1:4, and 1:5. A GFP control was used to visualize transfection efficiency at the four different lipofectamine: plasmid ratios. The lipofection mix was added to each respective 12-well plate containing 500 µL of media. The lipofected cells were incubated in a 5% CO_2 incubator for approximately 48 hours before visualizing GFP expression and extracting DNA for downstream analysis.

The C2C12 cells were maintained in culture medium supplemented with 20% Fetal Bovine Serum (FBS) and 1% Pen-Strep. The cells were cultured in two 15cm flasks, grown to 70% confluency. The cells were trypsinized for 2 minutes and 10 mL of culture medium was added to collect cells into a 15 mL tube. The cells were centrifuged at 250xg for 5 min and resuspended in 200 µL of PBS. Two washes were performed, with centrifugation at 250xg between each step. A PBS blank was prepared and the electroporator was adjusted to the following settings: Voltage (V) = 160V, Capacitance (C) = 1000, Resistance = Max. 20 nanograms of plasmid DNA was added to the cells, with an equal mass ratio of Cas9 plasmid to gRNA plasmid. Thus, 5 ug of $gRNA_1$ plasmid, 5 µg of $gRNA_2$ plasmid, and 10 µg of Cas9 was added to the cells. The cells were shocked and, moving quickly, 500 µL of culture medium was added to the sample. After pipetting 1-2 times, the sample was moved into a single well of a 6-well plate. The 6-well plate was moved to the CO2 incubator and incubated for 72 hours without disturbing.

After 72 hours of incubation, the electroporated cells were passaged onto a fresh 6 well plate in fresh growth medium and grown to 70-80% confluency. After growing to 70-80% confluency, the growth medium was replaced with differentiation medium consisting of Dulbecco's Modified Eagles Medium (DMEM), insulin, and donor equine serum. Differentiation media was added every 24 hours for 7 days before DNA extraction and preparation for sequencing.

2.3 Analysis

2.3.1 NIH3T3 Lipofection

Deletion PCR was used to assess the deletion efficiency of the dual guide lipofection. If the transfection was successful, the deleted fragment will drop out of the parent band and can be visualized on the gel. The primers were designed to amplify across the genomic deletion region. Lane 1 (L1) shows the control untreated sample, Lane 2 (L2) and Lane 3 (L3) are treated samples, at different lipofectamine concentrations, that show the 1638 base pair (bp) parent band and the 467 bp fragment with exon 23 deleted.

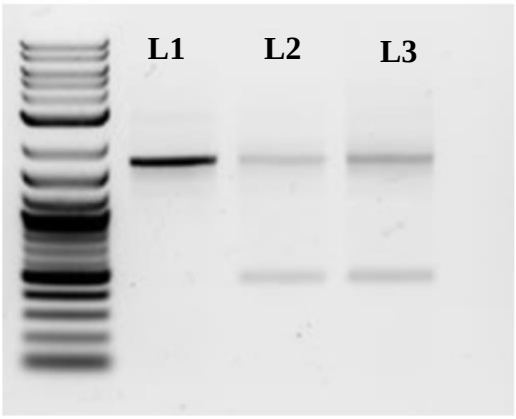

Figure 2.3: PCR gel showing untreated control (L1) and dual guide treated (L2 and L3) at different plasmid:lipoefectamine ratios. The top band reveals the parent band, with the bottom band in L2 and L3 revealing effective deletion of Exon 23.

2.3.2 C2C12 Electroporation

To confirm deletion of exon 23 at the cDNA level in C2C12 cell line, deletion PCR was performed using primers designed to flank exon 23 in the cDNA. Figure 2.4 below demonstrates the upper parent band, with an expected length of 467 bp and the deleted fragment below, missing exon 23. In Figure 2.4 below, there is a 430 bp parent band with the 213 bp deletion band shown below.

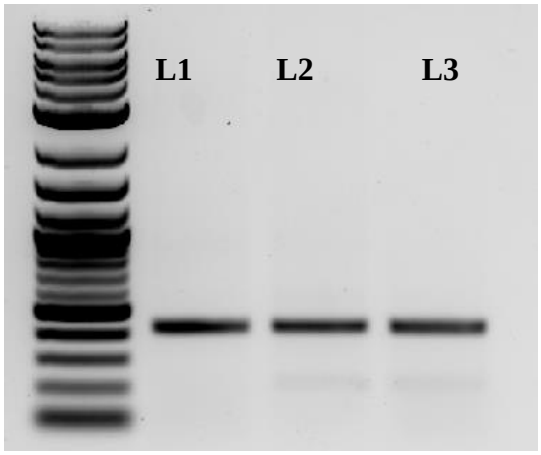

Figure 2.4: Deletion PCR showing untreated control (L1) compared to treated groups (L2 and L3) with deletion band shown

Chapter 3: Sample Preparation and Oxford Nanopore Sequencing

3.1 Introduction

Sample preparation involves the extraction of DNA and RNA and preparation of these samples for sequencing using the Oxford Nanopore MinION device. Both DNA and RNA are extracted for downstream sequencing analysis. As an initial effort, DNA was extracted from NIH3T3 cells and RNA was later extracted from dystrophin-expressing C2C12 cells for sequencing of cDNA.

As mentioned briefly in the first chapter, Oxford Nanopore uses a novel sequencing technique whereby DNA strands are passed through a protein nanopore. As the DNA passes through the nanopore, changes in electrical current are measured and subsequent basecalling algorithms assign a nucleotide base to each electrical current signature that is created. In this section, four total samples are run on the Oxford Nanopore device: treated and untreated DNA sample obtained from NIH3T3 cell line and treated and untreated cDNA sample from C2C12 cell line. The maximum 10 kb DNA amplicon is used, and a 3 kb cDNA amplicon is used, as this was the maximum PCR product that was able to be achieved.

3.2 Methods

3.2.1 Sample Preparation

DNA Extraction from NIH3T3 Cell Line

72-hour post-transfection, the cells were trypsinized and collected into a 15 mL tube. The sample was centrifuged at 250xg for 5 minutes. The cell pellet was resuspended in 200 µL of Phosphate Buffered Saline (PBS) and 20 µl of Proteinase K added. The DNA extraction was performed according to the DNeasy Blood & Tissue (Qiagen) kit protocol. Primers were designed using Primer 3 and NCBI Primer Design tool. Several different primer sets were ordered, including primers that produced a 3 kb, 5 kb, and 10 kb amplicon.

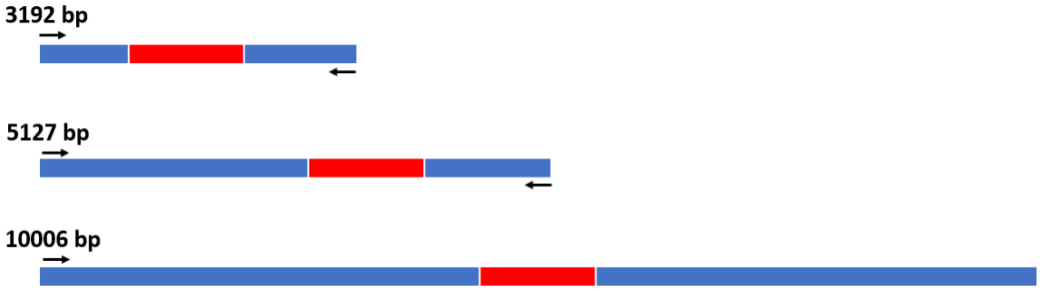

Figure 3.1: 3 kb, 5 kb, and 10 kb primers designed to flank the target region containing exon 23, with the target region being approximately central in the amplicon

There were several parameters that were taken into consideration when selecting primers. Primer pairs with a GC-content of approximately 55%, a primer length of approximately 20 bases, melting temperatures within 3 degrees Celsius, and minimal self-complementarity and self 3' complementarity. The primers were designed such that an amplicon would be produced with the target region approximately at the center of the amplicon product (shown in Figure 3.1).

RNA Extraction and cDNA Preparation from C2C12 Cell Line

RNA was extracted from C2C12 cell line after 7 days of differentiation. The sample was spun down at 250xg for 5 minutes, washed with PBS media, and the pellet stored in -80C until ready for extraction. The RNA was extracted using the RNeasy kit according to manufacturer protocol (Qiagen). VILO kit was used to produce cDNA, according to the manufacturer's instructions. The desired target region was PCR enriched, with primers optimized to produce a 3 kilobase (3 kb) amplicon.

Sample Preparation for Sequencing

Sample purification was performed, and the OD 260/280 and OD 230/60 were measured using the Nanodrop instrument to assess purity of the sample. The OD 260/280 was approximately 1.8 and the OD 230/260 was between 2.0-2.2, indicating a pure sample. Sample purity is important for sequencing as an impure sample can clog the nanopore flow cells, decreasing the overall sequencing yield of the run.

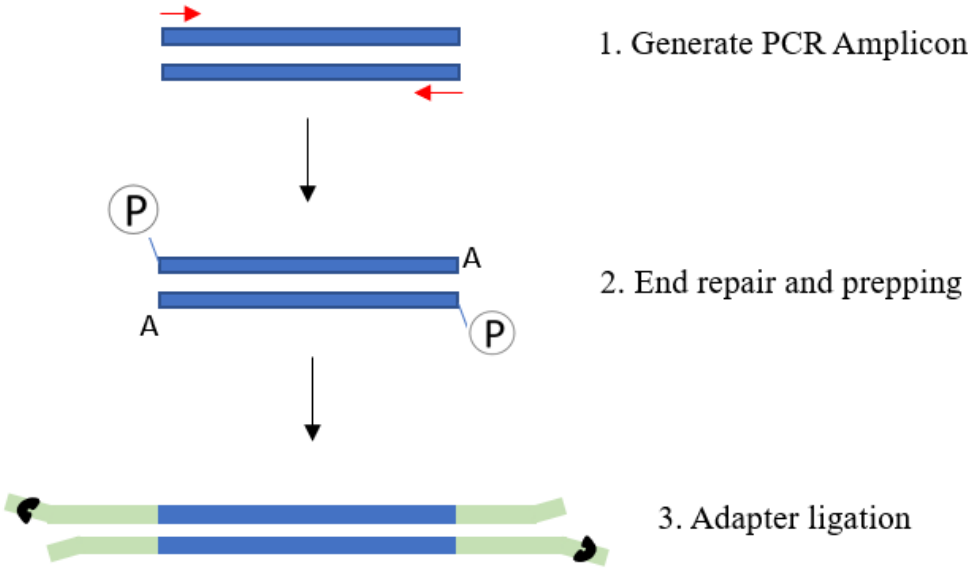

Figure 3.2: Library preparation including generation of a PCR amplicon, end repair and prepping, and adapter ligation

The DNA and cDNA samples were both prepared using the Ligation Sequencing Kit (Figure 3.2) from Oxford Nanopore kit in combination with the NEBNext Companion kit for end preparation. According to the sequencing preparation protocol, 100-200 fmol of DNA was added to a centrifuge tube and diluted with nuclease-free water to 47 µl volume. The following mixture was prepared in a PCR tube: 1 µl of DNA CS, 47 µl of DNA, 3.5 µl NEBNext FFPE DNA repair buffer, 2 µl NEBNext FFPE DNA Repair mix, 3.5 µl Ultra II End-prep reaction buffer, and 3 µl Ultra II End-prep enzyme mix. After mixing, the sample was incubated at 20 degrees C for 5 minutes and 65 degrees C for 5 minutes. After the incubation was complete, the mixture was cleaned up using the AMPure XP beads by transferring the DNA sample to a clean microcentrifuge tube and adding 60 µl, in a 1:1 volume ratio, of resuspended AMPure XP beads and mixing by gently flicking the tube. According to the Ligation Sequencing kit protocol, adapters were ligated to the DNA ends in preparation for sequencing.

The prepared DNA library was loaded onto the flow cell, and the DNA was sequenced over approximately 3 hours and data collected using the MinKNOW software. The sequencing run was performed using the default parameters. Live basecalling was performed by the MinKNOW software and FASTQ files for outputted to an Oxford Nanopore file on local workstation computer.

The structural variation pipeline (found on GitHub) was used for analysis of variants across the enriched and sequenced region of DNA. The pipeline maps reads using lra, estimates parameters for variant calling based on read depth, calls variants using cuteSV, and filters variants by minimum and maximum length, read support, and variation type and these default parameters are shown in Figure 3.3 (Jiang et al., 2020; Pedersen & Quinlan, 2018; Ren & Chaisson, 2020).

Minimum SV length	30
Maximum SV length	100000
Minimum read length	1000
Minimum read mapping quality	20

Figure 3.3: Filter parameters for structural variant pipeline

3.3 Results & Analysis

3.3.1 Enrichment

A treated and untreated 10 kilobase (kb) gDNA amplicon was generated (Figure 3.4) for sequencing and a treated and untreated 3 kb cDNA amplicon was generated. A 3 kb cDNA amplicon is used for sequencing because a 10 kb amplicon was unable to be obtained using PCR-based methods, potentially due to the fragmentation of the cDNA sample. The 3 kb amplicon

spans 18 exons, surrounding the Exon 23 target site, and will provide an initial framework for analyzing sequencing events at the cDNA level. Future work will address the size limitation of cDNA amplicons using PCR-based methods and other PCR-free methods will hopefully address this limitation.

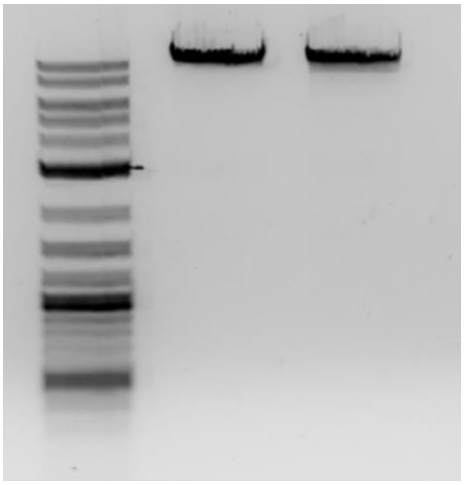

Figure 3.4: 10 kb gDNA amplicon showing untreated sample (L1) and treated sample (L2)

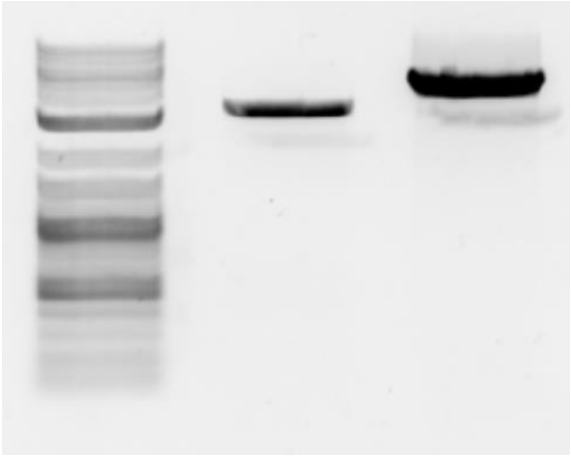

Figure 3.5: 3 kb gDNA amplicon (L1) and 5 kb gDNA amplicon (L2)

3.3.2 Quality Analysis

The basic Quality Control (QC) pipeline in Epi2Me labs was used to assess the quality of each sequencing run. The pipeline included a distribution of total reads, mean read length, N50, and mean Q scores. The pipeline also plotted a distribution of Q Scores, with a Q score greater than 7 being classified as a passed read. The pipeline also generated a distribution map showing the number of reads across the flow cell.

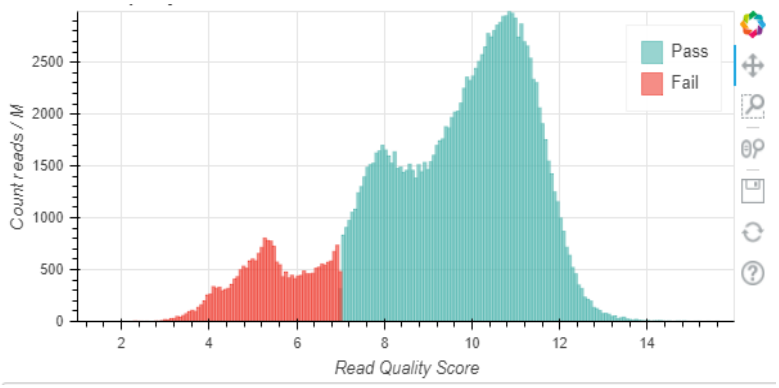

Figure 3.6: Read quality distribution

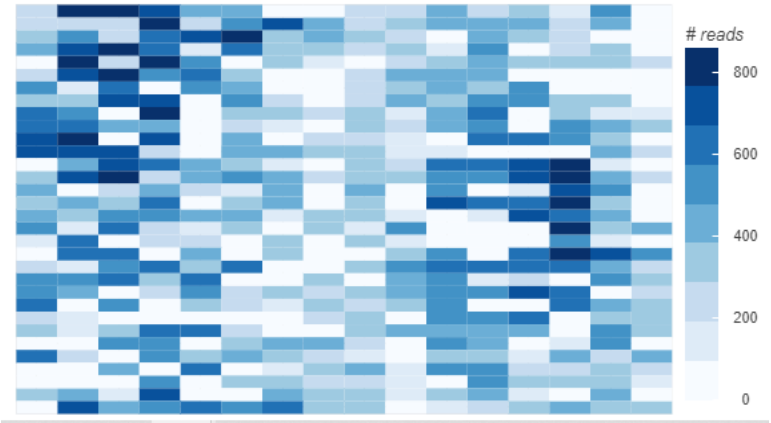

Figure 3.7: Read distribution across flow cell

3.3.3 DNA Sequencing of Treated and Untreated Samples

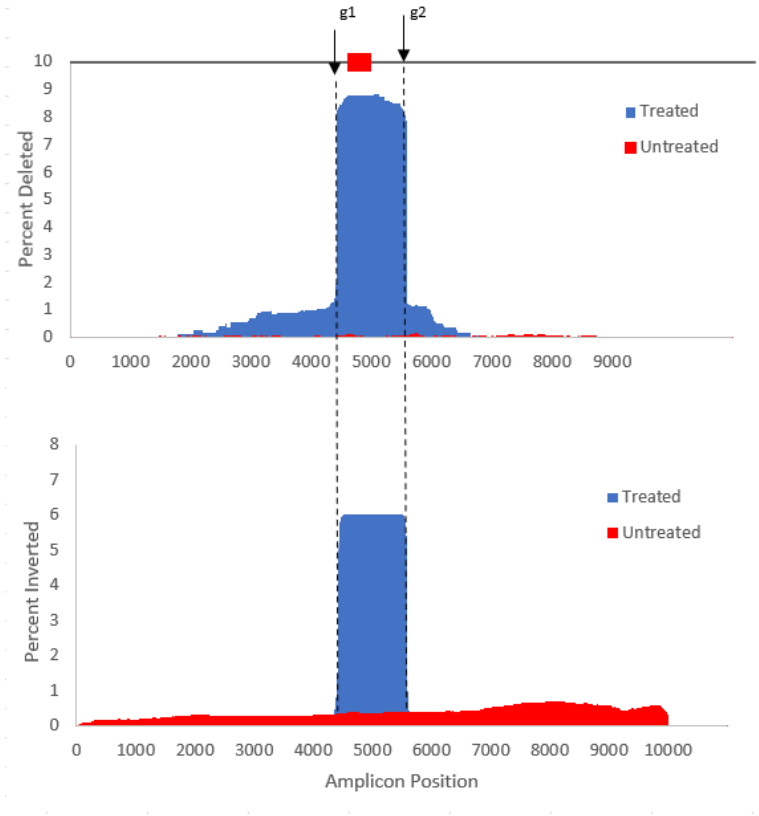

Figure 3.8: Genomic DNA (gDNA) sequencing run of treated and untreated samples showing percent deletion (top) and insertion (bottom) with exon 23 deletion being the most prevalent editing event

A 10 kilobase amplicon was optimized for both the untreated and treated samples, as shown in Figure 3.8, and sequencing reveals exon 23 deletion as the most prevalent editing event. Figure 3.9 reveals 8% deletion frequency for the targeted Exon 23 deletion, with other large deletions, up to several kilobases, that were detected around the targeted region. Notably, there is a higher off-target deletion frequency on the left side of the targeted region with an approximately 3.4 kilobase deletion spanning exons 22 and 23 as well as portions of intron 21 and 23. In contrast, the inversions were precise, with a 6% frequency. At the gDNA level, the majority of editing events are upstream of the 5' gRNA. This is potentially explained by increased microhomology or repetitive sequences on the left side of the target region. This could also be explained by

batch-to-batch variability and would have to be confirmed by increasing sample size. Future work could also focus on addressing this issue by looking into predictive algorithms that could explain this observation.

3.3.4 cDNA Sequencing of Treated and Untreated Samples

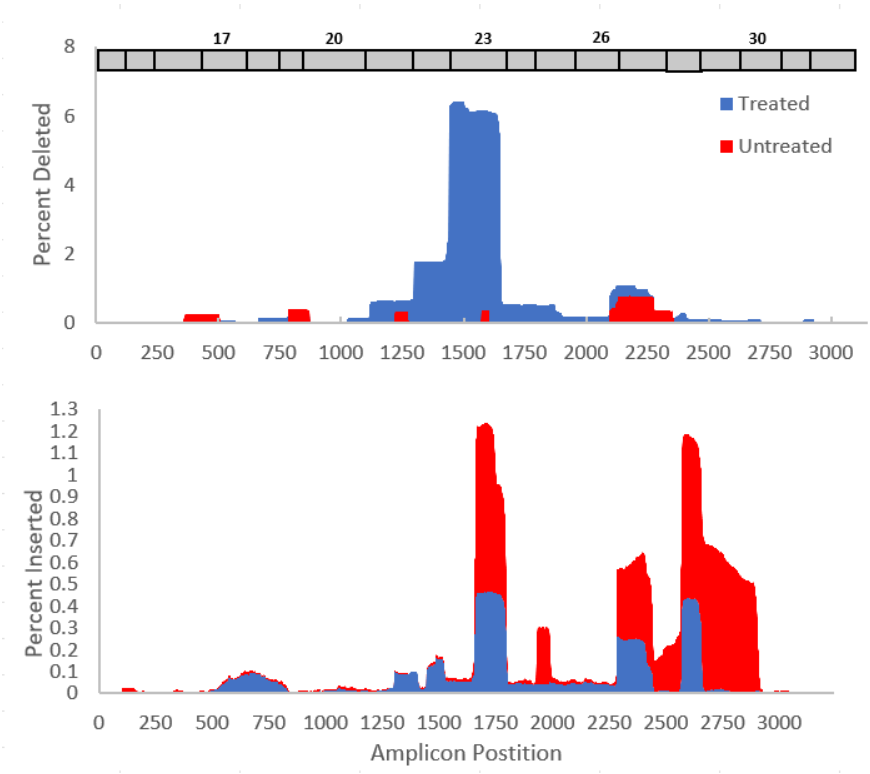

Figure 3.9: cDNA sequencing of 3 kb amplicon showing percent deletion (top) and insertion (bottom) with 3 kb deletion being most prevalent editing event

A 3.2 kilobase amplicon from the cDNA product was prepared spanning across 14 exons and exon 23 deletion was shown to be the most prevalent editing event, as shown in Figure 3.9. Multiple-exon deletions and other exon deletions were also detected. Notably, exon 27 deletion is detected in both the treated and untreated cDNA samples. While it is challenging to speculate whether this observation is reproducible, it could be explained by an artifact unique to C2C12s as well as an alignment error.

Interestingly, insertions in the untreated sample are similar or higher in frequency relative to the treated sample, indicating isoform diversity that is independent of treatment. 1-2% insertion is consistent with previous in vivo data showing 1-2% frequency of intron inclusion in untreated samples (Nelson et al., 2019). Insertions into the transcript include pseudoexons, alternate splice donor, and a duplication. These pseudoexons were located between canonical splice acceptor (5'-AG) and splice donor (GT-3'). Figure 4.0, shown below, depicts a 110 bp insertion, aligning to intron 21, and inserting at the 5' splice acceptor site.

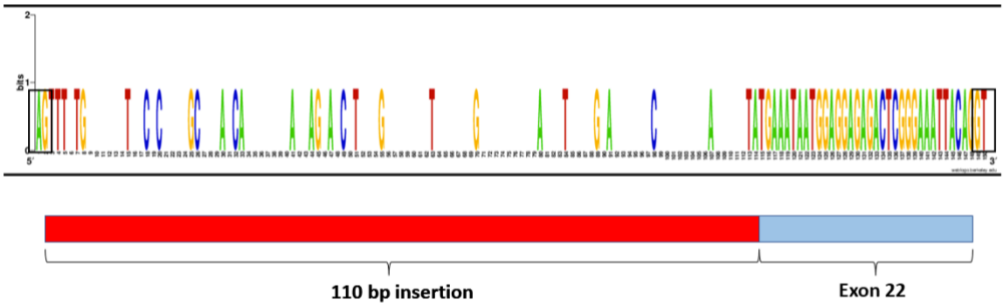

Figure 4.0: Depiction of 110 bp insertion into exon 22 with canonical AG-GT splice sites shown

No transcript reads aligned to the complementary strand indicating no detectable product from the inverted DNA. This may indicate that inversions contribute to productive gene editing in a similar fashion as shown for Leber congenital amaurosis type 10 (LCA10) (Maeder et al., 2019). To confirm this assertion, we could make a cell line containing the inversion and demonstrate normal dystrophin expression. These sequencing results establish the utility of long-read sequencing in assessing on-target editing efficiency as well as unintended off-target edits.

Chapter 4: Conclusion and Future Perspectives

4.1 Conclusion and Limitations

This work affirms the usefulness of long-read sequencing in assessing the on-target efficiency of CRISPR-Cas9 gene editing and further characterizing on-target precision. Notably, this technique can be used to optimize CRISPR-based therapies developed in the lab. This work provides an initial framework for characterizing CRISPR-Cas9 gene editing and future optimization of this technique will further improve read depth as well as amplicon length. Furthermore, a reliable technique for assessing gene editing efficiency is needed as CRISPR therapies move toward clinical translation. This technique is best paired orthogonally with other next-generation sequencing-based approaches. There are inherent limitations associated with using Oxford Nanopore sequencing, including lower read accuracy and ability to detect single nucleotide variations (SNV) (Kono & Arakawa, 2019). Continued improvements in pore chemistry and base-calling algorithms increase the reliability of nanopore sequencing, but an Illumina-based approached overcomes the limitations associated with the current status of nanopore sequencing technology.

4.2 Future Perspectives

The scope of this project was to develop a technique for assessing on-target and off-target editing efficiencies of a dual-guide CRISPR-Cas9 system in vitro. While an in vitro study provided an initial basis for developing this technique, future studies will focus on applying the same technique to in vivo models of Muscular Dystrophy. These samples were treated with the same dual-guide system targeting exon 23, and we expect to see vector integration in addition to many of the cryptic events observed in vitro (Hanlon et al., 2019). The sample size of this project was also insufficient to conclusively characterize the structural variants identified and increasing the

number of samples will verify the reproducibility of structural variants observed. Optimization of this protocol was focused on in vitro samples, but future work will focus on in vivo treated and untreated samples. In order to further optimize this technique, additional enrichment approaches such as CRISPR-Cas9 and hybridization-based approaches should be explored. These approaches mitigate the amplification bias that is inherent in using PCR for enrichment. Other enrichment approaches will also enable the sequencing of much larger DNA and cDNA amplicons, allowing for a more comprehensive assessment of unintended genomic alterations. In this work, the maximum amplicons for DNA and cDNA were 10 kilobases and 3 kilobases, respectively. The editing events observed with this sequencing events highlights the needs for potential alternatives. It is well-characterized that the double-stranded breaks induced by the non-homologous end joining mechanism results in undesired outcomes. Moreover, many disease states result from specific INDELS that require a more precise editing technique (Anzalone et al., 2019). Thus, our lab is also exploring prime editing as an alternative to the double-stranded break induced by CRISPR-based strategies. Additional measures to detect sample preparation need to be implemented to further optimize the quality of the run. In this study, 260/280 and 230/260 ratios were observed to assess DNA purity. Additionally, the nanodrop instrument was used to approximate DNA quantity because it was readily available in the laboratory. In future sequencing runs, a Qubit instrument should be used to quantify more precisely 100-200 fmol input onto the flow cell. An optimized technique for effectively analyzing the precision and efficacy of editing mechanisms will guide future research efforts in the Nelson Lab. Much of the work in the lab is moving toward delivery strategies for CRISPR gene editing, including non-viral delivery vectors. Additionally, prime editing as a more precise alternative compared to other CRISPR-based methods is being explored (Chemello et al., 2021a).

References

Albert, T. J., Molla, M. N., Muzny, D. M., Nazareth, L., Wheeler, D., Song, X., Richmond, T. A., Middle, C. M., Rodesch, M. J., Packard, C. J., Weinstock, G. M., & Gibbs, R. A. (2007). *Direct selection of human genomic loci by microarray hybridization*. https://doi.org/10.1038/NMETH1111

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology 2020 21:1*, *21*(1), 1–16. https://doi.org/10.1186/S13059-020-1935-5

Amoasii, L., Hildyard, J. C. W., Li, H., Sanchez-Ortiz, E., Mireault, A., Caballero, D., Harron, R., Stathopoulou, T. R., Massey, C., Shelton, J. M., Bassel-Duby, R., Piercy, R. J., & Olson, E. N. (2018). Gene editing restores dystrophin expression in a canine model of Duchenne muscular dystrophy. *Science*, *362*(6410), 86–91. https://doi.org/10.1126/SCIENCE.AAU1549/SUPPL_FILE/AAU1549_AMOASII_SM.PDF

Anzalone, A. v, Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, *576*, 149. https://doi.org/10.1038/s41586-019-1711-4

Chamberlain, J. R., & Chamberlain, J. S. (2017). Progress toward Gene Therapy for Duchenne Muscular Dystrophy. In *Molecular Therapy* (Vol. 25, Issue 5, pp. 1125–1131). https://doi.org/10.1016/j.ymthe.2017.02.019

Chemello, F., Chai, A. C., Li, H., Rodriguez-Caycedo, C., Sanchez-Ortiz, E., Atmanli, A., Mireault, A. A., Liu, N., Bassel-Duby, R., & Olson, E. N. (2021a). Precise correction of Duchenne muscular dystrophy exon deletion mutations by base and prime editing. In *Sci. Adv* (Vol. 7). https://www.science.org

Chemello, F., Chai, A. C., Li, H., Rodriguez-Caycedo, C., Sanchez-Ortiz, E., Atmanli, A., Mireault, A. A., Liu, N., Bassel-Duby, R., & Olson, E. N. (2021b). Precise correction of Duchenne muscular dystrophy exon deletion mutations by base and prime editing. *Science Advances*, *7*(18). https://doi.org/10.1126/SCIADV.ABG4910/SUPPL_FILE/ABG4910_SM.PDF

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science (New York, N.Y.)*, *339*(6121), 819. https://doi.org/10.1126/SCIENCE.1231143

Dapprich, J., Ferriola, D., Mackiewicz, K., Clark, P. M., Rappaport, E., D'Arcy, M., Sasson, A., Gai, X., Schug, J., Kaestner, K. H., & Monos, D. (2016). The next generation of target capture technologies - large DNA fragment enrichment and sequencing determines regional

genomic variation of high complexity. *BMC Genomics*, *17*(1), 1–14. https://doi.org/10.1186/S12864-016-2836-6/TABLES/3

Duan, D. (2018). *Micro-Dystrophin Gene Therapy Goes Systemic in Duchenne Muscular Dystrophy Patients*. https://doi.org/10.1089/hum.2018.012

Fajrial, A. K., He, Q. Q., Wirusanti, N. I., Slansky, J. E., & Ding, X. (2020). A review of emerging physical transfection methods for CRISPR/Cas9-mediated gene editing. *Theranostics*, *10*(12), 5532. https://doi.org/10.7150/THNO.43465

Gao, Q., & McNally, E. M. (2015). The Dystrophin Complex: structure, function and implications for therapy. *Comprehensive Physiology*, *5*(3), 1223. https://doi.org/10.1002/CPHY.C140048

Giannoukos, G., Ciulla, D. M., Marco, E., Abdulkerim, H. S., Barrera, L. A., Bothmer, A., Dhanapal, V., Gloskowski, S. W., Jayaram, H., Maeder, M. L., Skor, M. N., Wang, T., Myer, V. E., & Wilson, C. J. (2018). UDiTaS^{TM}, A genome editing detection method for indels and genome rearrangements. *BMC Genomics*, *19*(1), 1–10. https://doi.org/10.1186/S12864-018-4561-9/TABLES/4

Gilpatrick, T., Lee, I., Graham, J. E., Raimondeau, E., Bowen, R., Heron, A., Downs, B., Sukumar, S., Sedlazeck, F. J., & Timp, W. (2020). Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nature Biotechnology 2020 38:4*, *38*(4), 433–438. https://doi.org/10.1038/s41587-020-0407-5

Givens, B. E., Naguib, Y. W., Geary, S. M., Devor, E. J., & Salem, A. K. (2018). *Nanoparticle-Based Delivery of CRISPR/Cas9 Genome-Editing Therapeutics*. https://doi.org/10.1208/s12248-018-0267-9

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., Leproust, E. M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D. B., Lander, E. S., & Nusbaum, C. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *NATURE BIOTECHNOLOGY*, *27*. https://doi.org/10.1038/nbt.1523

Hanlon, K. S., Kleinstiver, B. P., Garcia, S. P., Zaborowski, M. P., Volak, A., Spirig, S. E., Muller, A., Sousa, A. A., Tsai, S. Q., Bengtsson, N. E., Lööv, C., Ingelsson, M., Chamberlain, J. S., Corey, D. P., Aryee, M. J., Joung, J. K., Breakefield, X. O., Maguire, C. A., & György, B. (2019). *High levels of AAV vector integration into CRISPR-induced DNA breaks*. https://doi.org/10.1038/s41467-019-12449-2

Hendel, A., Fine, E. J., Bao, G., & Porteus, M. H. (2015). Quantifying On and Off-Target Genome Editing. *Trends Biotechnol*, *33*(2), 132–140. https://doi.org/10.1016/j.tibtech.2014.12.001

Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*, *157*(6), 1262. https://doi.org/10.1016/J.CELL.2014.05.010

Ishino, Y., Krupovic, M., & Forterre, P. (2018). History of CRISPR-Cas from encounter with a mysterious repeated sequence to genome editing technology. *Journal of Bacteriology*,

200(7). https://doi.org/10.1128/JB.00580-17/ASSET/DE8FEBC5-B202-4B94-A69A-7B1049774EFA/ASSETS/GRAPHIC/ZJB9990946770007.JPEG

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakatura, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of Bacteriology*, *169*(12), 5429–5433. https://doi.org/10.1128/JB.169.12.5429-5433.1987

Jia, H., Guo, Y., Zhao, W., & Wang, K. (2014). Long-range PCR in next-generation sequencing: comparison of six enzymes and evaluation on the MiSeq sequencer. *Scientific Reports 2014 4:1*, *4*(1), 1–8. https://doi.org/10.1038/srep05737

Jiang, T., Liu, Y., Jiang, Y., Li, J., Gao, Y., Cui, Z., Liu, Y., Liu, B., & Wang, Y. (2020). Long-read-based human genomic structural variation detection with cuteSV. *Genome Biology*, *21*(1), 1–24. https://doi.org/10.1186/S13059-020-02107-Y/FIGURES/5

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). *A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity*. https://www.science.org

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., & Doudna, J. (2013). *RNA-programmed genome editing in human cells*. *2*, 471. https://doi.org/10.7554/eLife.00471

Kono, N., & Arakawa, K. (2019). Nanopore sequencing: Review of potential applications in functional genomics. *Development, Growth & Differentiation*, *61*(5), 316–326. https://doi.org/10.1111/DGD.12608

Lim, K. R. Q., Maruyama, R., & Yokota, T. (2017). Eteplirsen in the treatment of Duchenne muscular dystrophy. *Drug Design, Development and Therapy*, *11*, 533. https://doi.org/10.2147/DDDT.S97635

Lim, K. R. Q., Yoon, C., & Yokota, T. (2018). Applications of CRISPR/Cas9 for the Treatment of Duchenne Muscular Dystrophy. *Journal of Personalized Medicine*, *8*(4). https://doi.org/10.3390/JPM8040038

Logsdon, G. A., Vollger, M. R., & Eichler, E. E. (2020). Long-read human genome sequencing and its applications. *Nature Reviews. Genetics*, *21*(10), 597. https://doi.org/10.1038/S41576-020-0236-X

Mackenzie, S. J., Nicolau, S., Connolly, A. M., & Mendell, J. R. (2021). Therapeutic Approaches for Duchenne Muscular Dystrophy: Old and New. *Seminars in Pediatric Neurology*, *37*, 100877. https://doi.org/10.1016/J.SPEN.2021.100877

Maeder, M. L., Stefanidakis, M., Wilson, C. J., Baral, R., Barrera, L. A., Bounoutas, G. S., Bumcrot, D., Chao, H., Ciulla, D. M., Dasilva, J. A., Dass, A., Dhanapal, V., Fennell, T. J., Friedland, A. E., Giannoukos, G., Gloskowski, S. W., Glucksmann, A., Gotta, G. M., Jayaram, H., … Jiang, H. (2019). Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. *Nature Medicine*, *25*, 229–233. https://doi.org/10.1038/s41591-018-0327-9

Mann, C. J., Honeyman, K., Cheng, A. J., Ly, T., Lloyd, F., Fletcher, S., Morgan, J. E., Partridge, T. A., & Wilton, S. D. (2001). *Antisense-induced exon skipping and synthesis of dystrophin in the mdx mouse*. www.pnas.orgcgidoi10.1073pnas.011408598

Mendell, J. R., Sahenk, Z., Lehman, K., Nease, C., Lowes, L. P., Miller, N. F., Iammarino, M. A., Alfano, L. N., Nicholl, A., Al-Zaidy, S., Lewis, S., Church, K., Shell, R., Cripe, L. H., Potter, R. A., Griffin, D. A., Pozsgai, E., Dugar, A., Hogan, M., & Rodino-Klapac, L. R. (2020). Assessment of Systemic Delivery of rAAVrh74.MHCK7.micro-dystrophin in Children With Duchenne Muscular Dystrophy: A Nonrandomized Controlled Trial. *JAMA Neurology*, *77*(9), 1122–1131. https://doi.org/10.1001/JAMANEUROL.2020.1484

Mojica, F. J. M., Ce´sar Dıéz-Villasenõr, C., Jesu´s Garcıá-Martı´nezmartı´nez, J., & Soria, E. (2005). *Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements*. https://doi.org/10.1007/s00239-004-0046-3

Nakamura, A. (2019). Mutation-Based Therapeutic Strategies for Duchenne Muscular Dystrophy: From Genetic Diagnosis to Therapy. *Journal of Personalized Medicine*, *9*(1), 1–21. https://doi.org/10.3390/JPM9010016

Nelson, C. E., Hakim, C. H., Ousterout, D. G., Thakore, P. I., Moreb, E. A., Castellanos Rivera, R. M., Madhavan, S., Pan, X., Ran, F. A., Yan, W. X., Asokan, A., Zhang, F., Duan, D., & Gersbach, C. A. (2016). In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science*, *351*(6271), 403–407. https://doi.org/10.1126/SCIENCE.AAD5143

Nelson, C. E., Wu, Y., Gemberling, M. P., Oliver, M. L., Waller, M. A., Bohning, J. D., Robinson-Hamm, J. N., Bulaklak, K., Castellanos Rivera, R. M., Collier, J. H., Asokan, A., & Gersbach, C. A. (2019). Long-term evaluation of AAV-CRISPR genome editing for Duchenne muscular dystrophy. *Nature Medicine 2019 25:3*, *25*(3), 427–432. https://doi.org/10.1038/s41591-019-0344-3

Olson, E. N. (2019). *Toward the correction of muscular dystrophy by gene editing*. https://doi.org/10.1073/pnas.2004840117

Olson, E. N. (2021). Toward the correction of muscular dystrophy by gene editing. *Proceedings of the National Academy of Sciences*, *118*(22). https://doi.org/10.1073/PNAS.2004840117

Pedersen, B. S., & Quinlan, A. R. (2018). Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics*, *34*(5), 867–868. https://doi.org/10.1093/BIOINFORMATICS/BTX699

Ren, J., & Chaisson, M. J. (2020). lra: the Long Read Aligner for Sequences and Contigs. *BioRxiv*, 2020.11.15.383273. https://doi.org/10.1101/2020.11.15.383273

Schneider, J., Gonzalez, J., Brown, K., Golebiowski, D., Ricotti, V., Quiroz, J., & Morris, C. (2017). SGT-001 Micro-dystrophin gene therapy for Duchenne muscular dystrophy. *Neuromuscular Disorders*, *27*, S188. https://doi.org/10.1016/J.NMD.2017.06.343

Schultzhaus, Z., Wang, Z., & Stenger, D. (2021). CRISPR-based enrichment strategies for targeted sequencing. *Biotechnology Advances*, *46*, 107672. https://doi.org/10.1016/J.BIOTECHADV.2020.107672

Shin, J. H., Pan, X., Hakim, C. H., Yang, H. T., Yue, Y., Zhang, K., Terjung, R. L., & Duan, D. (2013). Microdystrophin Ameliorates Muscular Dystrophy in the Canine Model of Duchenne Muscular Dystrophy. *Molecular Therapy*, *21*(4), 750–757. https://doi.org/10.1038/MT.2012.283

Tabebordbar, M., Zhu, K., Cheng, J. K. W., Chew, W. L., Widrick, J. J., Yan, W. X., Maesner, C., Wu, E. Y., Xiao, R., Ran, F. A., Cong, L., Zhang, F., Vandenberghe, L. H., Church, G. M., & Wagers, A. J. (2016). In vivo gene editing in dystrophic mouse muscle and muscle stem cells. *Science (New York, N.Y.)*, *351*(6271), 407. https://doi.org/10.1126/SCIENCE.AAD5177

Tremblay, J. P., Iyombe-Engembe, J.-P., Duchêne, B., & Ouellet, D. L. (2016). Gene Editing for Duchenne Muscular Dystrophy Using the CRISPR/Cas9 Technology: The Importance of Fine-tuning the Approach. *Molecular Therapy*, *24*(11), 1888–1889. https://doi.org/10.1038/MT.2016.191

Wang, B., Li, J., & Xiao, X. (2000). Adeno-associated virus vector carrying human minidystrophin genes effectively ameliorates muscular dystrophy in mdx mouse model. *Proceedings of the National Academy of Sciences*, *97*(25), 13714–13719. https://doi.org/10.1073/PNAS.240335297

Weirather, J. L., Cesare, M. de, Wang, Y., Piazza, P., Sebastiano, V., Wang, X.-J., Buck, D., & Au, K. F. (2017). Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Research*, *6*. https://doi.org/10.12688/F1000RESEARCH.10571.2

Xiao-Jie, L., Hui-Ying, X., Zun-Ping, K., Jin-Lian, C., Li-Juan, J., & Chen, J.-L. (2015). *CRISPR-Cas9: a new and promising player in gene therapy*. https://doi.org/10.1136/jmedgenet-2014-102968

You, L., Tong, R., Li, M., Liu, Y., Xue, J., & Lu, Y. (2019). Advancements and Obstacles of CRISPR-Cas9 Technology in Translational Research. *Molecular Therapy - Methods & Clinical Development*, *13*, 359–370. https://doi.org/10.1016/J.OMTM.2019.02.008

Yue, Y., Ghosh, A., Long, C., Bostick, B., Smith, B. F., Kornegay, J. N., & Duan, D. (2008). A Single Intravenous Injection of Adeno-associated Virus Serotype-9 Leads to Whole Body Skeletal Muscle Transduction in Dogs. *Molecular Therapy*, *16*(12), 1944–1952. https://doi.org/10.1038/MT.2008.207

Zhang, Y., Li, H., Min, Y.-L., Sanchez-Ortiz, E., Huang, J., Mireault, A. A., Shelton, J. M., Kim, J., Mammen, P. A., Bassel-Duby, R., & Olson, E. N. (2020). Enhanced CRISPR-Cas9 correction of Duchenne muscular dystrophy in mice by a self-complementary AAV delivery system. In *Sci. Adv* (Vol. 6). http://advances.sciencemag.org/

Appendix: PCR Primer Design

gDNA Primers	Forward	Reverse
3 kb	ATGGAGGAGAGACTCGGGAA	AATGGACAAGCACCTCCCAC
5 kb	AAGGCCAATCTCCATTTCTGGT	AACAGTCCCTTCAGTTGGGG
10 kb	ACAACTGCCTCTGTCACTTCTC	TGCAAATCACCACATGTTCCAT

cDNA Primers	Forward	Reverse
3 kb	GAGCAGGTCAGGGTCAACTC	ACCTTCGCACCCAACTCATT
5 kb	TAGCTGGTCCGACGGGTTG	AGGAGATCATCAGCCTGCCTCTT
10 kb	GTGTTGGGCTCACTCACTTGC	GTAGTGCATTTTATGGCCCTTTGC

Deletion PCR primers	Forward	Reverse
gDNA	TTTGTTGATTCTAAAAATCCCATGT	GGACTGAAGAACTTGGAGAAGGA
cDNA	GGATCCAGCAGTCAGAAAGC	TCACCAACTAAAAGTCTGCATTG