SPLICING ANALYSIS OF A MYOTONIC DYSTROPHY TYPE 1 MOUSE MODEL TREATED WITH SMALL MOLECULES

by

NATALIE A. PELLITIER

A THESIS

Presented to the Department of Biology
and the Robert D. Clark Honors College
in partial fulfillment of the requirements for the degree of
Bachelor of Science

June 2017

An Abstract of the Thesis of

Natalie A. Pellitier for the degree of Bachelor of Science
in the Department of Biology to be taken June 2017

Title: Splicing Analysis of a Myotonic Dystrophy Type 1 Mouse Model Treated with Small Molecules

Approved: _____

Michael J. Harms

Myotonic dystrophy type 1 (DM1) is the most common form of adult-onset muscular dystrophy which has a wide variety of symptoms, but is known for its classic symptom myotonia. Myotonia is the ability to contract, but not relax one's muscles. The mis-splicing of a chloride channel gene *Clcn1* causes myotonia in DM1 patients. To address symptoms such as myotonia, this thesis aims to test the effect of two potential treatments, heptamidine and furamidine, on splicing events in a DM1 mouse model.

The HSA^{LR} mouse model replicates the tri-nucleotide CTG repeat that DM1 patients have in the 3' UTR of the DMPK gene. Repeats in affected individuals range from 38 to more than thousands. This transgenic mouse expresses 220 CUG repeats which form a toxic RNA structure. Like a sponge, the toxic RNA soaks up the MBNL proteins. MBNL proteins are responsible for regulating the alternative splicing of fetal vs. adult isoforms of more than 200 splicing events. When sequestered, the MBNL proteins can no longer regulate these events, resulting in mis-splicing of its targets (such as *Clcn1*), leading to the symptoms of the disease (like myotonia).

Two small molecules have been shown to rescue (reverse) mis-splicing in the transgenic mouse model. Heptamidine and furamidine were each administered to three sibling-matched mice. Three additional mice, which also expressed the CUG repeats, did not receive drug treatment as a control. Their transcriptomes were sequenced and a MISO (Mixture-of-Isoforms) splicing analysis was preformed to quantify and visualize splicing rescue across treatments. Gene expression changes were visualized to compare treatment off-target effects. This quantitation and comparison will confirm and identify known splicing biomarkers of DM1, leading to an improved understanding of how these small molecules are acting at a molecular level. This knowledge will inform future drug design to treat DM1 and other diseases caused by toxic repeat RNA.

Acknowledgements

First and foremost, I would like to thank Professor Berglund and Dr. Leslie Coonrod for your patience in teaching me and for investing time, energy, and resources into my development as a scientist and my understanding of biomedical research. Thank you Dr. B for your endless positivity and for always pouring love and wisdom into my personal and professional development. Thank you, Dr. Harms for advising me as a primary advisor outside of your specialty, and for your guidance in discussing the intersectionality of science and religion. Thank you, Hailey Olafson for always responding to my panicked texts regarding errors in my scripts and for being a wonderful friend. And thank you to Stacey Wagner for your encouragement and guidance regarding MISO. I'd also like to thank Dr. Subramony who graciously taught me about neuromuscular disorders through shadowing in his clinic. Finally, I'd also like to thank my family and close friends who constantly support me, inspire me, and have provided me with endless encouragement to complete this project.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Myotonic dystrophy type 1 (DM1), the most common form of adult-onset muscular dystrophy, does not yet have a cure. Individuals with DM1 have difficulty with daily tasks such as picking up a mug or grabbing a door knob, because they cannot seamlessly relax their muscles and release from these objects. This is the classic symptom of DM1: myotonia. Yet for many patients with myotonic dystrophy, myotonia is not their biggest complaint, instead it's gastrointestinal dysfunction, regurgitation, and dyspepsia which interfere with daily life [3]. Interestingly, while it's not reported as frequently in the literature, the 20-something DM1 patients I saw while shadowing in a muscular dystrophy clinic reported fatigue and sleepiness as being the most frustrating symptom. Thus, many take prescribed stimulants or self-medicated via copious amounts of caffeine, which can have their own effects on one's body. DM1 itself is a multisystem disorder "affecting skeletal muscle as well as eyes, heart, gastrointestinal tract, endocrine system, and central nervous system" [2].

While there are two forms of myotonic dystrophy with similar molecular mechanisms, DM1and DM2, DM1 most commonly affects adolescents and young adults. With a prevalence of 1 in 8000 individuals, DM1 is twice as common as two well-known diseases: Hodgkin Lymphoma and Huntington's disease [2]. There are three clinical diagnoses denoting the severity of an individual's symptoms: mild, classic, and congenital. While symptoms manifest on a spectrum, classic DM1 symptoms include cataracts, myotonia, and cardiac conduction abnormalities and reduced life span, whereas patients with congenital myotonic dystrophy have early childhood onset of severe symptoms [4, 15]. Adult-onset myotonic dystrophy symptoms

vary greatly between patients, and depend largely on the extent of the disease-causing genetic mutation. Complications due to muscle wasting in both heart and lungs lead to early patient mortality, most commonly due to respiratory muscle failure and cardiac abnormalities [17]. Symptoms also increase in severity with successive generations, and as an adult onset disease, most individuals are not aware they have the disease before they have children. Due to its autosomal dominant nature, children of a parent with DM1 have a 50 percent chance of inheriting the disease-causing mutation, which can cause feelings of guilt. Additionally, one-third of DM1 patients are diagnosed with depression in the early stages of disease onset, accompanied by increased levels of apathy, reduced initiative, and fatigue [22].

With the goal of improving treatment available to patients with DM1, this project uses bioinformatics, or computational biology, to explore two potential treatments: small-molecules heptamidine and furamidine. While this thesis focuses on the splicing analysis of a DM1 mouse model treated with small molecules, addressing it without integrating the social impact seems shallow.

In combination with finding a treatment for patients with DM1, another issue arises: the difficulty of writing about biomedical research with empathy and understanding of patient's experiences. Many emotions can come with diagnoses of neuromuscular diseases that can generate fear, apprehension, and guilt which are not caused by the disease-causing mutation. Muscle strength also begins to weaken, leading to decreased ability to perform daily tasks. I intend to capture the weight of this disease as I discuss process of my thesis. In writing about diseases like myotonic dystrophy, we

must wrestle with the reality that these genetic diseases affect much more than pure gene expression, affecting quality of life above all else.

I gained a more holistic understanding of myotonic dystrophy by getting outside of the lab through shadowing Dr. Subramony in his muscular dystrophy clinic in July 2016. I had the opportunity to understand patient afflictions through their visit with a specialist, which may have been an initial or follow-up visit. While there are the occasional drug trials that patients can participate in, for most DM patients the current care plan included symptom tracking and management. Overall, these patients were actively invested in the DM scientific and patient communities, and surprised me with their resiliency and general positivity. Most patients had come to terms with their disorder, had undergone genetic testing, and were looking forward to maintaining as much muscular strength as possible.

Speaking with patients at the community research symposium also affirmed the importance of this project. While these potential therapeutics may not be released individually, an understanding of their mechanisms of action will provide knowledge for informed future drug design for use with patients.

BACKGROUND

Myotonic Dystrophy Physiology and Molecular Mechanism

At the molecular level, DM1 is caused by a three-nucleotide CUG repeat in the RNA of the 3' untranslated region of *the dystrophia myotonica protein kinase (DMPK)* gene on chromosome 19, shown in Figure 1 [27].

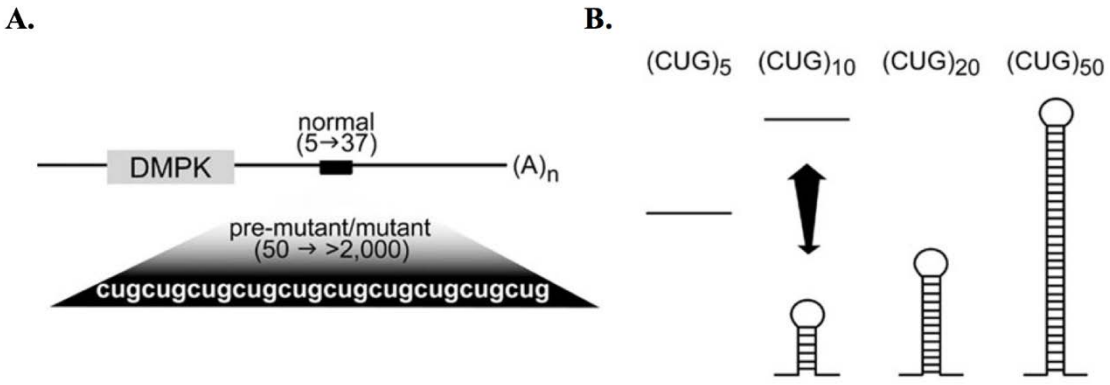

Figure 1. The location of DM1 expansion mutation in A) DNA and B) RNA gain-of-function structures. These structures lead to the sequestration of MBNL proteins. Figure was adapted from Figure 1 of [13].

While an unaffected individual may have anywhere from 5-37 CUG repeats, an individual with DM1 may have 38-4000 CUG repeats, with the number of repeats correlating with disease severity [10, 24]. This toxic RNA structure soaks up muscleblind-like proteins (MBNL) like a sponge absorbs water, due to binding sites on MBNL proteins for the CUG-repeats [20]. MBNL proteins are a family of proteins that regulate alternative splicing and determine which version of the gene (fetal or adult) is expressed in a cell at a given time [15]. Toxic RNA also forms nuclear foci, which concentrate in the nucleus and provide a high density of MBNL binding sites and a cellular environment where MBNL activity is lost [18]. When MBNL is sequestered to this toxic RNA structure and retained in the nucleus, it cannot regulate alternative

splicing elsewhere (Figure 2). Thus, many gene targets of MBNL are mis-spliced, leading to the variety of symptoms.

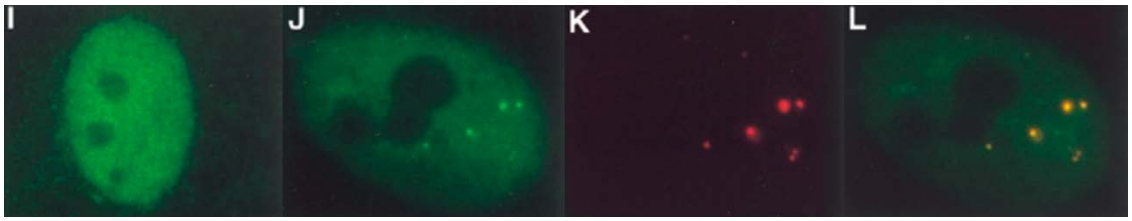

Figure 2. Pathogenic mechanism of nuclear foci in DM1 fibroblast cells. Figure adapted from [7]. Visualization of co-localization of MBNL and nuclear foci in the nucleus of DM1 cells by overexpression of green fluorescent protein (GFP)-tagged MBNL, and immunofluorescence. I) control fibroblast cells which were transfected with MBNL-GFP. J) DM1 fibroblast cell transfected with MBNL-GFP. K) DM1 fibroblast cell with CUG-repeat RNA labeled in red. L) Co-localization of MBNL and CUG-repeat containing nuclear foci are shown, indicating that MBNL aggregates on the CUG repeats in the DMPK repeat containing transcript, limiting the supply of free MBNL for alternative splicing regulation.

Understanding Splicing

Cells have developed mechanisms for processing genetic material to turn on and off protein expression in response to environmental cues. Splicing, which removes non-protein coding parts of DNA, is one such mechanism. Some of these non-protein coding portions of the DNA are called introns, while the transcribed portions of the DNA are called exons. To generate the proper set of proteins in the cell (given specific environmental conditions and other cell signaling) these pre-mRNA transcripts must be processed, or spliced (Figure 3A). However, as each cell contains the organism's entire DNA, some exons will be necessary in some cells and not in others: this process is referred to an alternative splicing (Figure 3B).

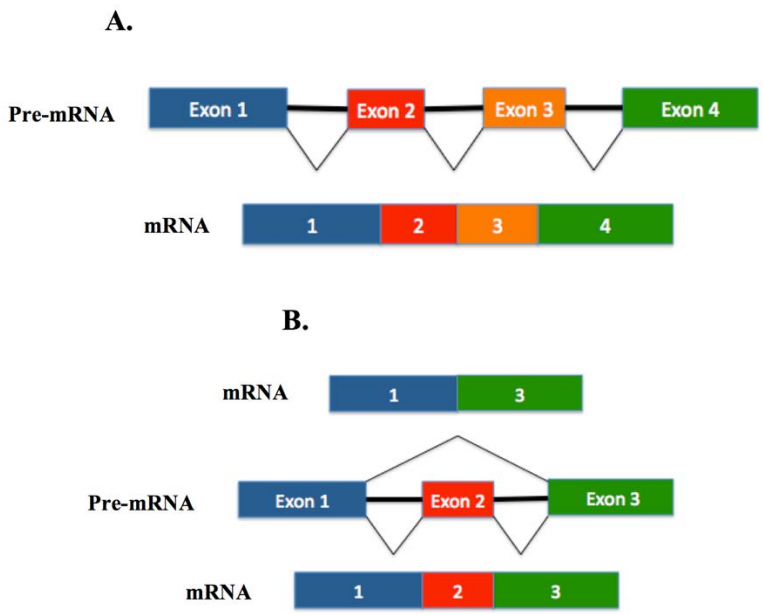

Figure 3. Splicing is carried out on the newly transcribed RNA. Not all parts of the RNA are protein coding, thus the exons are spliced together and the introns are removed. The exons code for protein, and when pre-mRNA is fully processed into mature RNA (mRNA), it can be translated into proteins. All exons can be included (A), or exons can be skipped (B). Skipped exons, or cassette exons, are specific sequences that are required in some proteins, but not others (depending on the cell type, developmental stage, etc.). In this example, exon 2 can be either included or excluded in the processed mRNA, depending on what signals the cell is receiving.

If proteins that regulate alternative splicing are sequestered and cannot regulate this process, then improper gene forms are expressed and disease symptoms displayed. In patients with DM1, the set of mis-spliced genes includes *Clcn1*, the mis-splicing of which directly results in the symptom myotonia. Muscle weakness has also been correlated with numerous splicing defects. Thus, addressing mis-splicing could directly affect symptoms of myotonic dystrophy [15].

A Clinical Perspective

While a cure for DM1 is currently non-existent, there are some therapies that patients with DM1 can access to help alleviate symptoms. To determine a patient care plan, physicians refer patients to neurologists when they suspect there is a neurological component to their disease. Once referred, patients may see a neuromuscular specialist such as Dr. S. Subramony, whom I shadowed at the University of Florida. Then physicians typically do a neurological assessment to diagnose, which may lead to genetic testing, the ability to sign up for drug trials, or physical tools such as orthotics to accommodate their worsening symptoms. They may receive various occupational therapies, such as hand training to counteract the myotonia, physical therapy, orthotics, and mobility therapy [1]. Neuromuscular specialists may also determine if patient could benefit from such therapies due to their range of motion and presence of residual muscle strength, a proxy for the presence of substantial muscle fibers [1]. While functional training, like hand therapy, will not cure symptoms, it contributes to an individual's abilities to live as independently and functionally as possible [1].

Other treatment is more medicalized. A review by Thornton *et al.* gives an overview of the current approaches to DM1 therapy. There are four mechanistic approaches including transcriptional silencing, post-transcriptional silencing, inhibiting interactions between MBNL and toxic RNA, and targeting pathways downstream of RNA toxicity [18]. While these approaches have yet to be fully-cleared for patient use, erythromycin, an antibiotic, reduces toxic RNA most successfully in comparison to 20 FDA-approved small molecules [14]. With no toxicity to humans, oral erythromycin

could prove the most effective therapy by reversing splicing and improving myotonia [14].

Current research also includes genome editing via CRISPR-Cas9 and stem cell therapy that could restore tissue health. A preliminary study conducted in 2015 shows that inserting a poly-A signal (a signal for early transcriptional termination) before the pathogenic repeat in the DMPK gene inhibits production of toxic RNA [23]. While this method could provide targeted tissue repair, such as in muscles, it's unlikely to be widely feasible until there are methods to grow and edit patient-derived stem pluripotent stem cells in a timely and cost-effective manner. However, once stem-cell therapies become more and more integrated into treatment of diseases, this may serve as an effective treatment free of off-target effects. CRISPR-Cas9 has also been used recently as a therapeutic via genome editing of removing the DMPK CUG repeat itself [19]. Again, CRISPR/Cas9 genome editing may provide patients with another option; however, it may not fully repair cells in a patient, as it would be limited to specific sites. While in the early stages, genome editing may be the future of DM1 therapeutics as we move more towards personalized medicine.

Using Small Molecules to Disrupt MBNL1 Interactions with CUG Repeats

Small molecules have been studied as potential therapeutics for treating DM1. Initially Warf *et al.* analyzed an FDA-approved anti-parasitic pentamidine, known to bind the minor groove of DNA, and they found that it could reverse splicing in a DM1 cell culture model [21]. Analogs were developed by varying the linker length in the middle of the molecule, and those were tested for toxicity and splicing rescue in a DM1 model [21]. This yielded heptamidine, which was found to rescue splicing through RT-PCR analysis [6]. Furamidine, another small-molecule, has also been tested on genetically modified mice. Heptamidine, one of the small molecules analyzed in this study, was previously shown rescue of splicing of two genes: *Clcn1* and *Atp2a1* (Figure 4). The HSA^{LR} transgenic mouse line was developed in 2000 to mimic DM1 symptoms with a CUG repeat length of 220 [12]. These HSA^{LR} mice have myotonia, and maintain approximately 81% of the alternative splicing consistent with DM1 patients, allowing for study of the splicing rescue of potential treatments [6, 15].

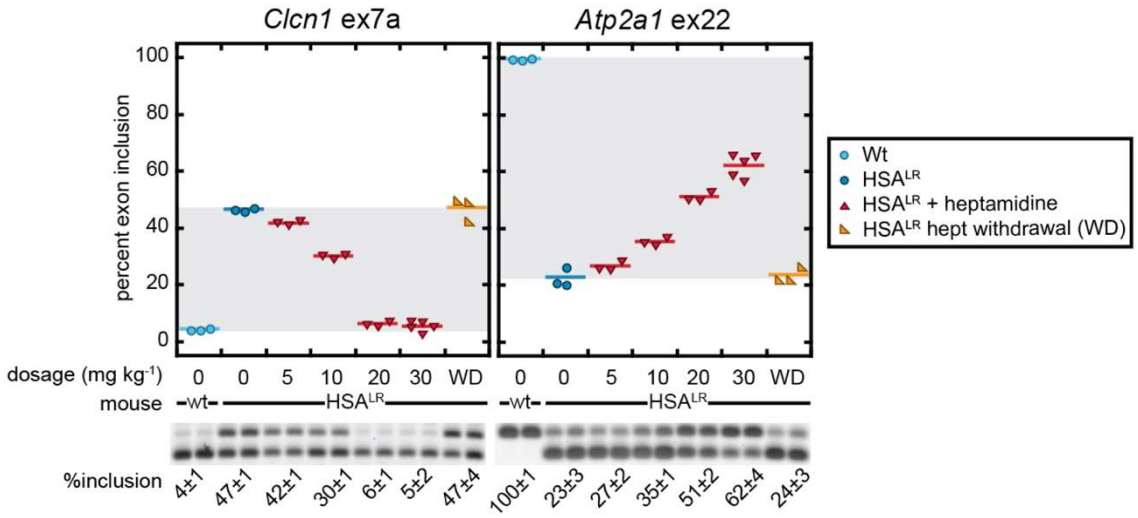

Figure 4. Dose-dependent rescue of mis-splicing of *Clcn1* and *Atp2a1* in HSA^{LR} mice treated with heptamidine. Data was recorded across a range of dosage of 0-30 mg kg^{-1} of heptamidine administered for a seven-day course. WD represents withdrawal state of 10-days post-treatment. An increase in dosage to 20 mg kg^{-1} heptamidine decreased the percentage of exon inclusion of *Clcn1* exon 7a to WT levels. In exon 22 of *Atp2a1*, percent exon inclusion directly correlated with increased heptamidine dosage reaching a peak of 60% inclusion with a dosage of 30 mg kg^{-1} [6].

This mis-splicing rescue demonstrates promising results of a potential treatment for patients with DM1. Heptamidine was administered to transgenic mice with a CUG_{220} repeat. Previous literature shows *Atp2a1* mis-splicing was not completely rescued by heptamidine, but was significantly improved. In the *Clcn1* gene, MBNL regulates the exclusion of exon 7a. When MBNL is sequestered by the CUG repeats, the mis-splicing in *Clcn1* leads to chloride channelopathy (dysfunction of the chlorine ion channel), and causes the classic symptom myotonia. Furamidine has also been shown to rescue mis-splicing of Clcn1 exon 7a and Atp2a1 exon 22 using qRT-PCR (Figure 5) [16].

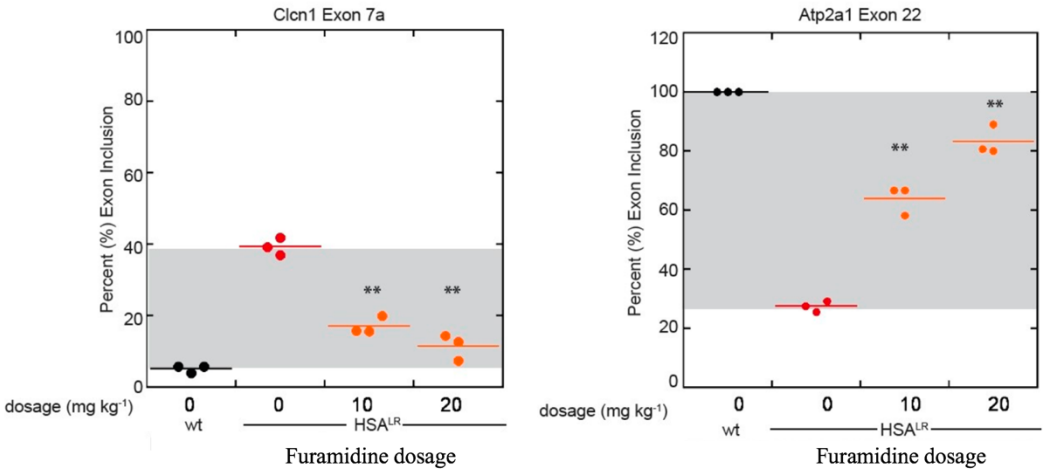

Figure 5. Dose-dependent rescue of mis-splicing of *Clcn1* and *Atp2a1* in HSA^{LR} mice treated with furamidine. Data was recorded across a range of dosage of 0, 10, and 20 mg kg^{-1} of furamidine administered over a seven-day course. WD represents withdrawal state of 10-days post-treatment. An increase in dosage to 20 mg kg^{-1} furamidine decreased the percentage of exon inclusion of *Clcn1* exon 7a approximately WT levels. In exon 22 of *Atp2a1*, percent exon inclusion directly correlated with increased furamidine dosage reaching a peak of upwards of 80% inclusion with a dosage of 20 mg kg^{-1} [16].

Heptamidine treatment also improves the symptom of myotonia in mice (Figure 6). In both genes, mis-splicing returned during withdrawal state, indicating the drug is not permanently correcting the disease.

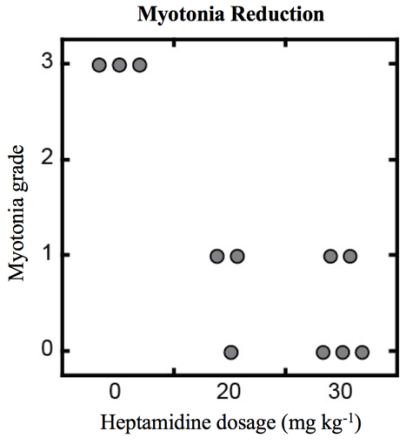

Figure 6. Myotonia reduction upon treatment with heptamidine. Treatment yields a significant reduction in myotonia [6].

While heptamidine may not cure DM1, it provides hope and knowledge for alleviating symptoms and halting the progression of muscle degeneration. A paper by Siboni *et al.* indicates furamidine acts on *Clcn1* and *Atp2a1* similarly, but with less toxicity. Our study is a continuation of the Coonrod *et al.* 2013 paper, with a narrowed focus on heptamidine and furamidine at a dosage of 30 mg kg^{-1} in hopes of further understanding the degree to which they affect splicing.

Heptamidine relieves myotonia, and furamidine, an FDA-approved anti-parasitic, also rescues the mis-splicing of *Clcn1* exon 7a (the mis-spliced event causing myotonia); however, what other mis-splicing events are they beneficially or adversely affecting? And what can we learn about their overall effects on gene expression? The splicing analysis conducted in this study will have two main focuses: to confirm that previously-studied mis-splicing of genes *Clcn1* and *Atp2a1* has been relieved, and to evaluate the efficacy of these treatments by comparing them to DM1 splicing biomarkers.

We will address these questions with a MISO analysis of RNA from transgenic mice. Through a gene expressions analysis and a splicing analysis, this study will yield further understanding of the way potential treatments function to alleviate patients of their debilitating symptoms. This thesis aims to combine current therapeutic research with a patient-focused understanding of the disease.

METHODS

Sibling matched HSA^{LR} mice were treated by Masayuki Nakamori at Osaka University. Sibling matching ensures that genetic differences among individuals are highly controlled, as siblings will have the most similar genomes apart from identical twins [6]. Nine mice were treated, three mice per treatment, with saline (control), 30-mg kg^{-1} furamidine, and 30-mg kg^{-1} heptamidine. After seven days of treatment, the tibialis anterior muscle was removed. RNA was extracted and shipped to the Berglund Lab where it was stored at -80 °C before it was processed.

To evaluate the state of RNA, the samples were sent to be analyzed using a fragment analyzer and showed partial degradation. This partial degradation could alter our analysis by affecting splicing coverage and gene expression, as certain genes or splicing junctions may have been stochastically degraded and thus not as well represented in the data. Regardless, we proceeded and adjusted post-sequencing to maintain adequate quality.

Protein-coding RNA, or messenger RNA (mRNA) is most important for our study, as the amounts of mRNA in a treated cell should vary among genes and show the molecular effects of furamidine and heptamidine. Since we cannot directly sequence RNA, we prepared DNA libraries from the RNA with the KAPA Stranded RNA-seq kit with Riboerase [30].

Library Preparation

Sequencing libraries (a collection of cDNA fragments to be sequenced), were prepared using the KAPA Stranded RNA-seq kit with Riboerase. The reasons for choosing this kit are two-fold: firstly, because our RNA was partially degraded, as

shown by the fragment analysis, a poly-A tail selection would be problematic, and secondly, per the recommendation of the sequencing facility and cost. The overview of the library preparation is shown in Figure 7.

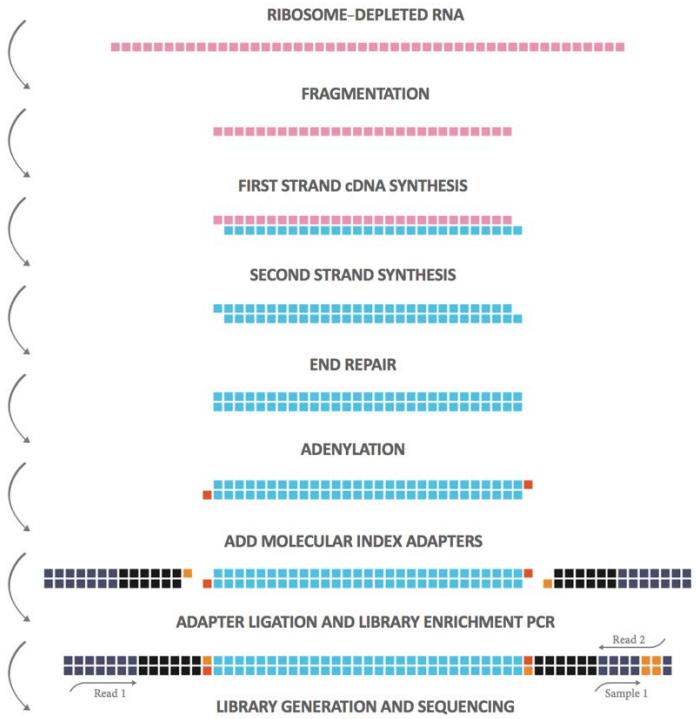

Figure 7. Schematic of Library Preparation. Adapted from the Bioo Scientific NEXT flex qRNA-seq RNA sample preparation which is also followed by the KAPA Stranded RNA-seq kit [25]. The workflow for the KAPA-stranded kit is similar [30].

Qubit Fluorometric Quantitation was used to measure the concentration of RNA to standardize the amount of RNA used across samples (Appendix 2). The KAPA-stranded kit procedure was followed apart from adding Bioo scientific molecular index adaptors in step 14 [20, 16].

Overview of Pipeline

Libraries were combined and sequenced on one lane using the NextSeq 500 sequencer in the University of Oregon's Genomics and Cell Characterization core facility. Using ACISS, the University of Oregon super computer, and HiperGator, the

University of Florida super computer, sequencing output was processed and analyzed using the workflow shown in Figure 8.

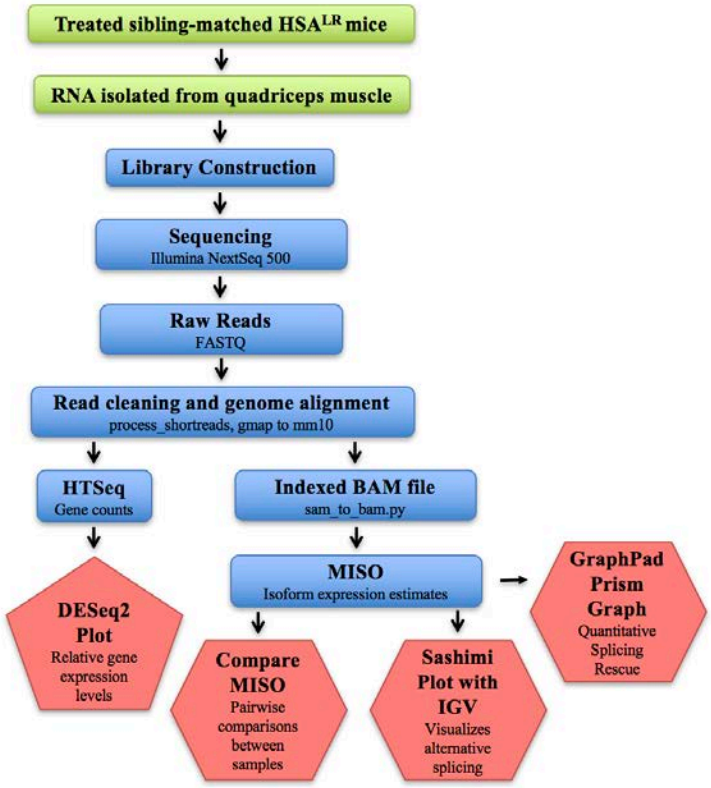

Figure 8. Workflow of the RNA-sequencing splicing analysis. Steps in green were conducted by M. Nakamori at Osaka University in 2015. Read cleaning, genome alignments, and HTSeq were run by Dr. Leslie Coonrod. The bam file was indexed myself, and I ran DESeq2, MISO, compare MISO, visualized alternative splicing with IGV, and represented alternative splicing with GraphPad Prism.

While three mice were treated with each drug (saline, furamidine, and heptamidine), only two heptamidine samples were analyzed as one sample, H-3, was removed from analysis due to its dis-similarity identified by an initial quality screening.

MISO stands for Mixture of Isoforms and it "quantitates the alternative splicing of genes from RNA-seq data" [27]. MISO uses known splice site information from a skipped exons annotations file (provided in their online documentation), to identify splice sites in RNA-seq data and splicing rescue across treatments. MISO

documentation was followed with minimal deviance to answer questions about the splicing rescue ability of these small molecules.

In pairwise comparisons between untreated and treated samples, MISO generates the estimate of the isoform (type of spliced gene) and the differential expression of isoforms for paired-end reads. MISO generates a Bayes factor for each event comparison, which is a quantitative statistical measure of differential expression of these isoforms [27]. These data were filtered with a minimal Bayes factor of 5, which is considered a substantial level of significance of the magnitude of splicing rescue, measured by present spliced in. These splicing events were compared to DM1 splicing biomarkers for less biased analysis, and to conclude more general effects on splicing.

To visualize differential gene expression, and understand off-target effects, the python script HTSeq was used to generate counts of gene transcripts, and DESeq2, an R package, was used to compare treated to untreated samples [29, 26]. Genes identified by DESeq2 were differentially expressed in the top 1% of expression changes affected by treatment, designated by a p-value of 0.01.

RESULTS

Heptamidine and Furamidine Rescue Splicing to Different Extents

Splicing rescue with treatment of furamidine and heptamidine can be both quantified and visualized to explain their effect and significance on the DM1 HSA^{LR} model. Initially, two previously-identified splice sites were analyzed, *Atp2a1* and *Clcn1* by Coonrod *et al.* Splicing, measured through PSI, percent exon inclusion, was quantified by MISO. Thirty-six splicing events were identified as mis-spliced in DM1, designated as 'splicing biomarkers' [15]. Splicing events that were significant with a Bayes factor of greater than five in both furamidine- and heptamidine-treated samples, that were also listed in the list of splicing biomarkers, are shown in Figure 9.

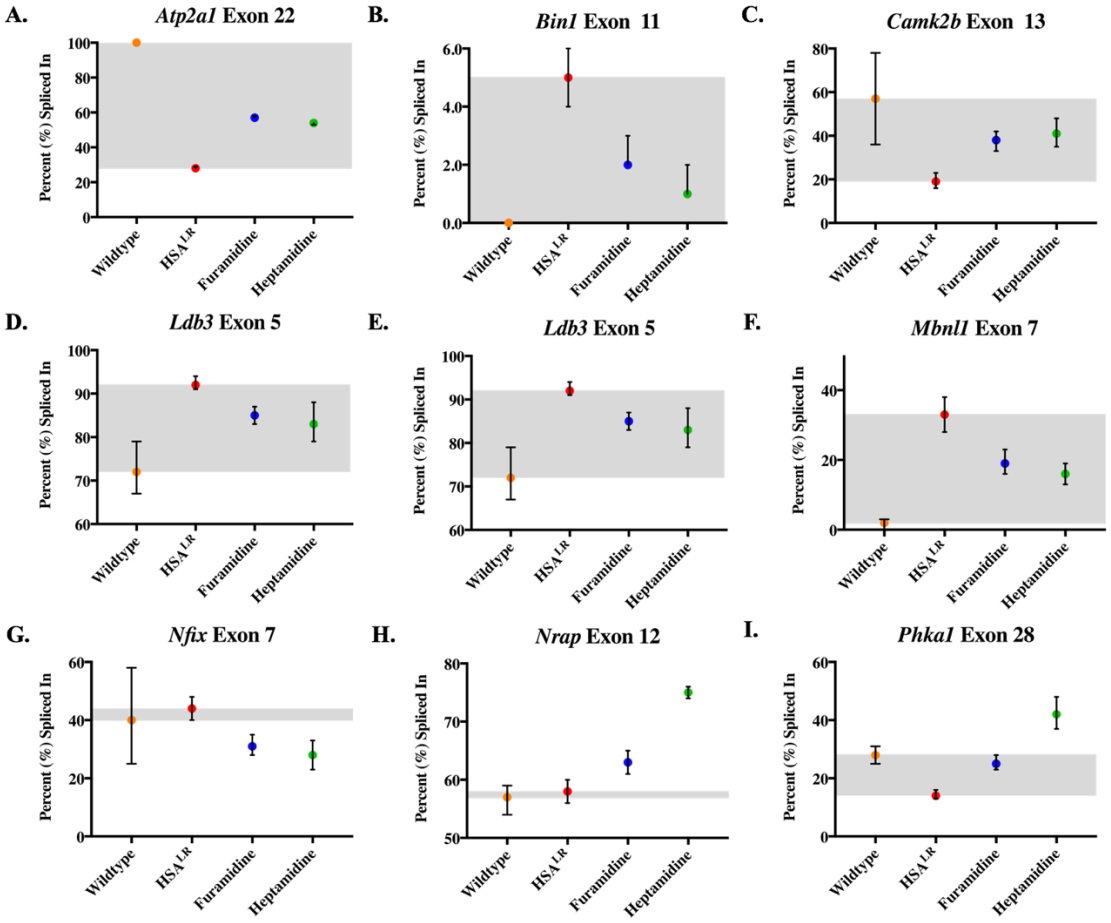

Figure 9. Percent Spliced In (PSI) across treatment in splicing biomarkers. Splicing biomarkers were identified in [15] and significant in both furamidine and heptamidine-treated samples with error bars representing confidence intervals identified by MISO. The grey shading shows the difference between percent exon inclusion across wildtype and HSA^{LR} mice. Splicing was partially rescued in these events apart from *Nrap* exon 12 where splicing was not rescued and was instead mis-spliced, and in *Nfix* exon 7 and *Phka1* exon 28 where splicing was over-corrected.

In the rescued splicing events shown in figure 8, 2/7 have better splicing rescue with furamidine treatment, while 5/7 events have better splicing rescue with heptamidine treatment.

The mis-splicing event causing myotonia, *Clcn1* exon 7a, shows a decrease in exon inclusion of exon 7a to wildtype-like levels. Wildtype mice have a mean exon inclusion of 10%, while control (disease) mice have inclusion of 30%. With

administration of furamidine exon inclusion levels decrease to 21% while they decrease to 13% with heptamidine treatment (Figure 10).

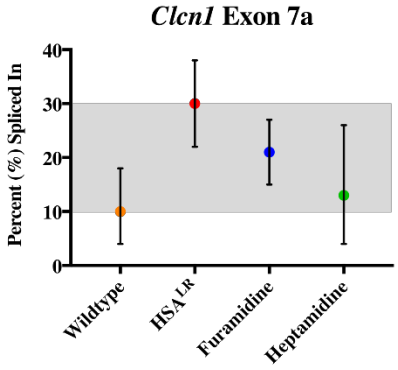

Figure 10. *Clcn1* exon 7a percent spliced in across treatments. Wildtype mice have about 10% inclusion of *Clcn1* exon 7a, while the disease-state mice have inclusion of 30%. Treated mice have partially-restored splicing of 21% inclusion with furamidine treatment and 13% exon inclusion with heptamidine treatment. Confidence intervals indicate that the mean percent exon inclusion changes are not significant (Bayes factor <5) between HSA^{LR} and furamidine, and HSA^{LR} and heptamidine.

Splicing rescue of *Atp2a1* exon 22 with the number of sequencing reads spanning the splicing junctions are shown (Figure 11); *Atp2a1* exon 22 is partially rescued upon treatment of heptamidine and furamidine (Figure 9A, 11).

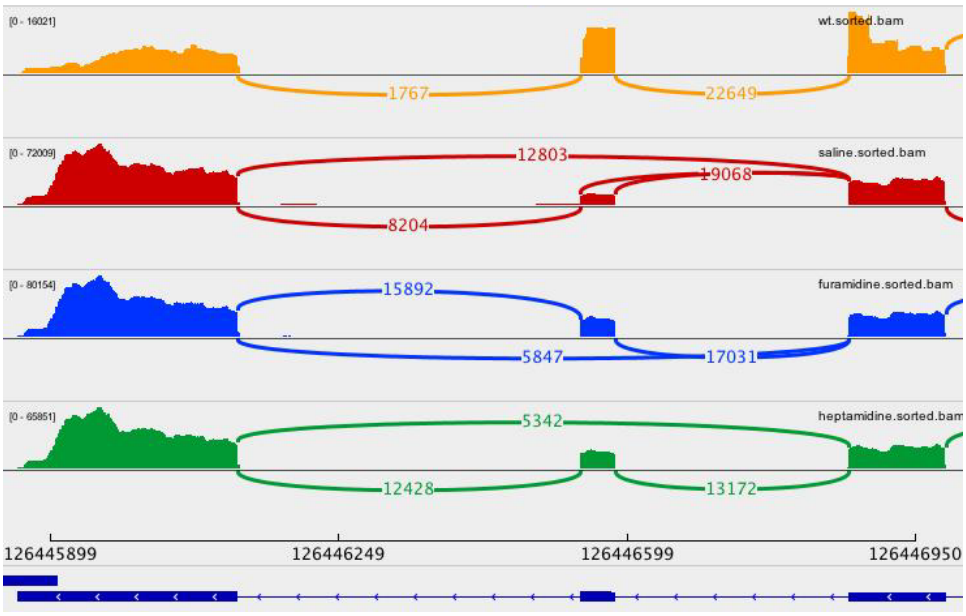

Figure 11. Splicing of gene *Atp2a1* exon 22 across treatments. In the wildtype, exon 22 is included. In the control sample (disease-state) exon 22 is excluded, shown in red, in comparison to the inclusion of exon 22 with treatment by furamidine and heptamidine which aligns more with wildtype-like exon inclusion.

The average number of reads spanning splicing junctions in *Atp2a1* exon 22 splicing is approximately 12,200 reads in comparison to an average of 17 reads across *Clcn1* exon 7a. While exon 7 splicing is corrected in a positive direction, the number of reads spanning the exon junction are not sufficient to draw significance as shown by the large confidence intervals in Figure 9 and the sashimi plot showing reads in Figure 12.

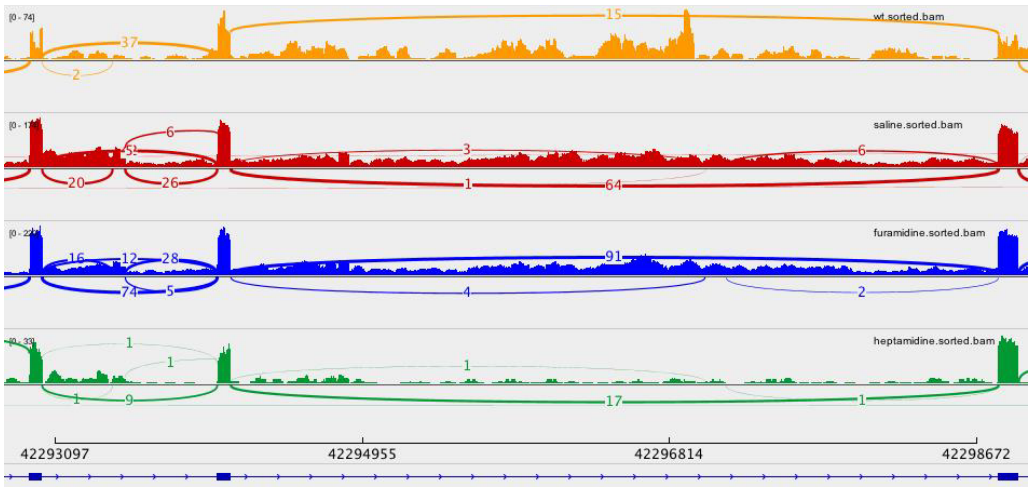

Figure 12. Splicing junction of *Clcn1* exon 7a across treatments. Mice treated with heptamidine have significantly better splicing rescue than those treated with furamidine although they are also rescued in a positive direction. Reads spanning the junction are minimal in all treatments.

Heptamidine and Furamidine Rescue Splicing of other DM1 biomarkers

Splicing events identified by MISO across treatments were gleaned for DM1 splicing biomarkers [15]. Of 42 splicing biomarkers, 9 were significantly affected by both heptamidine and furamidine treatment. Two events were significantly affected only by heptamidine, one was exclusively affected by furamidine, and 30 were either not found in our data or not significant (Figure 13). Full splicing biomarker PSI mean and confidence intervals across treatments are listed in Appendix 2, Table 1.

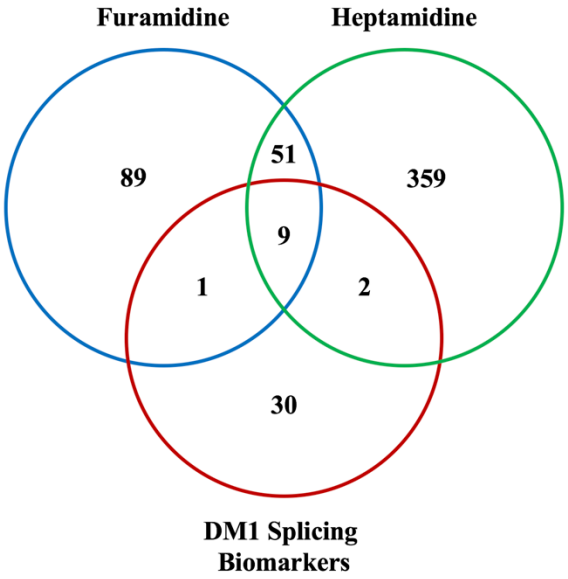

Figure 13. Splicing Rescue of DM1 splicing biomarkers by treatment. Forty-two splicing events were identified as splicing biomarkers of DM1 [15]. Of these, 9 were positively rescued by both heptamidine and furamidine, 2 were affected only by heptamidine, one was affected only by furamidine, and 30 were either not found in the data, or did not show significant splicing rescue.

Heptamidine and Furamidine Have Contrasting Effects on Gene Expression

Treatment with furamidine significantly affects gene expression less than treatment with heptamidine, indicating different mechanisms of action and off target effects (Figure 14). Genes that were significantly over- or under-expressed with a p-value of 0.01 or less in comparison to the control are shown in red, while the genes that are not significantly different are shown in grey. A p-value of 0.01 indicates that these genes are highly significant in their deviation from normal gene expression.

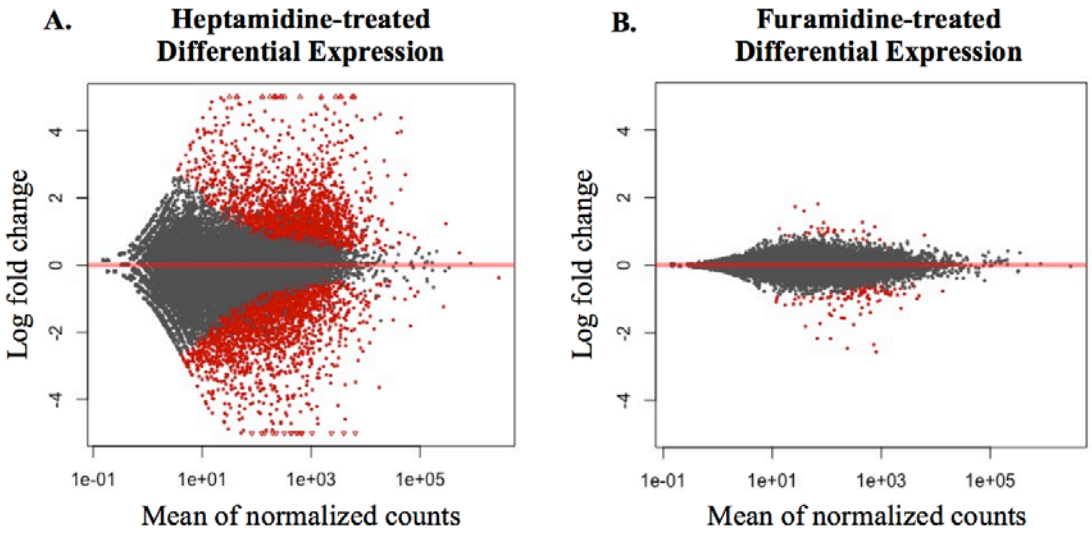

Figure 14. Differential gene expression with treatment of heptamidine and furamidine. Genes with differential expression to a p-value of less than 0.01 are shown in red. A) Treated by heptamidine 4465 genes were over-expressed (16.7%), and 4228 genes were under-expressed (15.7%). B) Treated with furamidine, 108 genes were over-expressed (0.39%), and 272 genes were under-expressed (0.99%).

DISCUSSION

Treatment of transgenic mice with small molecules heptamidine and furamidine rescue splicing of numerous mis-spliced events in DM1; however, the limited power of our study limits our ability to confirm effects observed in other studies. While both heptamidine and furamidine have been shown previously to rescue mis-splicing in a mouse model and HeLa cells, respectively, this RNA-seq analysis requires deeper sequencing. Figures 9-11 show the overall positive effects of heptamidine and furamidine on splicing. In comparison to DM1 splicing biomarkers causing disease symptoms, these small molecules do benefit transgenic mice, and thus could aid patients. However, their toxicity must be weighed with efficacy.

The gene expressions analysis, represented through DESeq2 plots demonstrate the off-target effects of both furamidine and heptamidine. As heptamidine is known to be toxic, the gene expression is consistent—where many genes are significantly affected by treatment. However, the gene expression plot of furamidine-treated mice show promise, as less than 1% of genes are differentially affected. This indicates that furamidine may be having few off-target effects and wouldn't interfere in other cellular processes.

The analysis of splicing using MISO showed splicing rescue of only two out of nine events previously studied in heptamidine. While the remaining 7/9 events trended positively, our study lacked the power to claim conclusive rescue (Bayes factor <5). Splicing effects were also considered with comparison to splicing biomarkers identified by Nakamori et. al [15]. Of these 36 events, nine of these events were significantly affected by furamidine and heptamidine. This may be due to the limitations of the

mouse model, the inadequate coverage of sequencing reads, or low gene expression levels. To get more coverage of splicing events, specifically *Clcn1* exon 7a, sequencing of the RNA libraries could be repeated to a greater depth, generating more reads and thus more reads spanning splicing junctions. The main barrier to addressing this limitation is the substantial monetary investment. Since other small molecules are being investigated, which show better efficacy and lower toxicity, there would be limited benefit to performing more sequencing of these data.

Limitations

Mouse models may not reveal all effects of treatment that could transpire in humans, for whom we intend to use the drugs. The HSA^{LR} model doesn't replicate heart, central nervous system, or muscle weakness in mice [9]. The HSA^{LR} model also does not include at least 12 of the human DM1 mis-spliced events [15]. These splicing markers would only be able to be studied in human tissue culture, and thus that could be the next step of this study. Furthermore, the mouse model also uses 220 repeats, which fits in the classic clinical diagnosis of DM1, which isn't extremely severe. Nakamori *et al.* showed a relationship between MBNL alternative splicing regulation and disease severity in human tissue, which would not be replicated through this model. They also suggest that "other splicing factors are also sequestered or indirectly affected through the activation of regeneration or stress response pathways" [15].

Future studies could address the number of replicate mice. As only three mice were treated with each drug there are few replicates, although this is rather maximal for sibling-matched mice. However, with a greater number of mice in the same generation of this transgenic mouse line, there may be ability to compare across numerous samples

to avoid the genomic differences that accompany different individuals. Regardless, to examine at splicing of three mice of each treatment, and two mice treated with heptamidine, these files were concatenated to run one ‘sample’ at once through MISO. This may have reduced the power of this study as well. Alternate software that can take replicates into account could be used to confirm these results and analyze the splicing across replicates: rMATS [28].

Despite limitations, our findings aid the development of new treatments and experiments to address symptoms of DM1 While the field of DM1 research is constantly evolving, this analysis fits into the niche of understanding the molecular basis for drug treatment and pre-clinical analysis of treatments.

CONCLUSION

Realistically, the likelihood of using heptamidine or furamidine for patients is slim; however, this analysis still adds to the body of knowledge of potential treatments. This analysis of splicing rescue could be used to develop new therapeutics by modifying their chemistry to be less toxic and to target specific cell types. If the scientific community better understands the exact mis-splicing in the disease, future drugs may be tailored to correcting specific mis-splicing events. Heptamidine and furamidine may also be combined with other drugs in a drug cocktail to be taken by patients, as they do generally positively affect DM1 mis-splicing events.

This analysis can be understood as an example of the many trial and errors of modern science. In the race to cure diseases, sometimes research yields imperfect treatments which propel future research, and other times we use toxic drugs out of desperation, for example with chemotherapy for cancers. However, in the case of DM1, the process of identifying a highly effective, FDA-approved, and non-toxic drug continues.

APPENDICES

Appendix 1. RNA-seq Library Preparation

The KAPA Stranded RNA-Seq Kit with RiboErase (HMR) kit was used for library preparation. The technical data sheet can be found at: https://www.kapabiosystems.com/product-applications/products/next-generation-sequencing-2/rna-library-preparation-2/kapa-stranded-rna-seq-riboerase/. Quantities were calculated for preparation of nine libraries, and the procedure was followed with minimal exceptions. In step 14.3, an annealing temperature of 65 °C was used rather than 60 °C in accordance with the enzyme specificity. In step 14, master mix was prepared using Next Flex Universal F primer, each sample was assigned a barcoded primer. Barcoded primer information can be accessed at: http://www.biooscientific.com/Portals/0/Manuals/NGS/5130-11-NEXTflex-qRNA-Seq-Kit-v2.pdf.

Appendix 2. Table 1

Table 1. Percent Exon Inclusion Across Treatments and Splicing Biomarkers

Splice event	WT PSI mean (low CI-high CI)	Control PSI mean (low CI-high CI)	Furamidine PSI mean (low CI-high CI)	Heptamidine PSI mean (low CI-high CI)	Effect of splicing change
ATP2A1 exon 22	100 (NA)	28 (28-29)	57 (57-58)*	54 (53-54)*	Ca++ reuptake
NFIX exon 7	40 (25-58)	44 (40-48)	31 (28-35)*	28 (23-33)*	transcription factor, myogenesis
INSR exon 11	13 (3-29)	17 (11-26)	24 (16-34)	23 (17-29)	insulin signaling (Δ signaling)
CAPZB exon 8	85 (77-91)	93 (90-95)	91 (89-94)	92 (89-94)	assembly of actin filaments
CACNA1S exon 29	40 (32-50)	87 (84-89)	94 (92-96)*	93 (90-96)	excitation contraction coupling (↑ Ca++ entry)
CAMK2B exon 13	57 (36-78)	19 (16-23)	38 (33-42)*	41 (35-48)*	Signaling
VPS39 exon 3	61 (15-96)	7 (1-21)	16 (7-35)	14 (3-35)	vesicle trafficking, TGFbeta signaling
CLCN1 exon 7a	10 (4-18)	30 (22-38)	21 (15-27)	13 (4-26)	chloride channel (loss of ion conductance)
LDB3 exon 11	6 (3-10)	68 (63-73	39 (36-43)*	24 (19-30)*	Z disc
GFPT1 exon 9	81 (42-100)	80 (67-90)	82 (71-91)	81 (68-92)	protein glycosylation (↓ feedback inhibition)
MBNL1 exon 7	2 (0-3)	33 (28-38	19 (16-23)*	16 (13-19)*	alternative splicing (↑ nuclear localization)
NRAP exon 12	57 (54-59)	58 (56-60)	63 (61-65)*	75 (74-76)*	myofibril assembly
ANK2 exon 21	89 (74-98)	53 (29-90)	62 (40-85)	69 (40-98)	membrane targeting
OPA1 exon 4b	30 (23-37)	28 (22-35)	38 (31-45)	43 (33-52)	mitochondrial dynamics
RYR1 exon 70	57 (51-62)	81 (79-83)	82 (80-84)	81 (79-83)	Ca++ release (↓ open probability)

* indicates significant changes with a Bayes factor >5 compared to HSA^{LR}

Splice event	WT PSI mean (low CI-high CI)	Control PSI mean (low CI-high CI)	Furamidine PSI mean (low CI-high CI)	Heptamidine PSI mean (low CI-high CI)	Effect of splicing change
COPZ2 exon 9b	7 (1-32)	4 (1-8)	7 (3-13)	3 (1-9)	vesicle trafficking
PHKA1 exon 28	28 (25-31)	14 (13-16)	25 (23-28)*	42 (37-48)*	muscle glycogenesis
FHOD1 exon 11a	33 (10-72)	34 (11-61)	33 (16-56)	47 (14-79)	actin organization
MBNL2 exon 7	15 (12-19)	21 (17-25)	15 (12-19)	31 (26-37)*	alternative splicing
DMD exon 78	97 (91-100)	99 (97-100)	99 (96-100)	99 (98-100)	membrane integrity
MLF1 exon 3	89 (85-92)	70 (64-76)	63 (57-68)	41 (32-52)*	oncoprotein
ABLIM2 exon 12	65 (35-88)	82 (62-97)	82 (61-98)	89 (68-99)	Z disc
BIN1 exon 11	0 (0-1)	5 (4-6)	2 (2-3)*	1 (1-2)*	T-tubule formation (↓ formation of T tubules)
LDB3 exon 5	72 (65-79)	92 (91-94)	85 (83-87)*	83 (79-88)*	Z disc
TXNL4A exon 4	5 (1-13)	2 (1-5)	3 (1-6)	2 (1-5)	PQBP1 binding protein

* indicates significant changes with a Bayes factor >5 compared to HSA^{LR}

Appendix 3. Table 2

Table 2. Library Preparation Information

Sample name	Treatment	µL RNA (408 ng)	µL H20	Qubit ng/µL	NextFlex Molecular Index Adapter
C-1	Saline	7.1	2.9	57.5	A1
C-2	Saline	7.17	2.83	56.9	A2
C-3	Saline	5.51	4.49	74	A3
H-1	30 mg/kg heptamidine	7.65	2.35	53.3	A4
H-2	30 mg/kg heptamidine	10	0	40.8	A5
H-3	30 mg/kg heptamidine	7.35	2.65	55.5	A6
F-1	30 mg/kg furamidine	7.11	2.89	57.4	A7
F-2	30 mg/kg furamidine	8.85	1.15	46.1	A8
F-3	30 mg/kg furamidine	7.27	2.73	56.1	A9

Appendix 3. MISO Pipeline run on the HiperGator command line

```
module load ufrc
srundev --time=4:00:00
module load python/2.7.6 samtools
python /home/eric.t.wang/Tools/pythonmodules/lib/python2.7/site-packages/misopy/
```

1. Example Sam to bam script in HiperGator:

```
python /ufrc/berglund/npellit3/2016_mouseRNAseq/sam_to_bam.py \
--convert /ufrc/berglund/npellit3/2016_mouseRNAseq/datanew/saline.sam \
/ufrc/berglund/npellit3/2016_mouseRNAseq/datanew/BAM \
--ref
/ufrc/berglund/npellit3/2016_mouseRNAseq/miso/Mus_musculus.GRCm38.68.dna.toplevel.fa.f
ai
```

2. Example MISO script in HiperGator:

```
python /home/eric.t.wang/Tools/pythonmodules/lib/python2.7/site-packages/misopy/miso.py \
--run /home/npellit3/2016_mouseRNAseq/miso/annotations/SE_indexed/ \
/home/npellit3/2016_mouseRNAseq/data2/Fur.sorted.bam \
--output-dir /home/npellit3/2016_mouseRNAseq/data2/Fur/ \
--read-len 66 \
--settings-filename /home/npellit3/2016_mouseRNAseq/miso/settings.txt \
--use-cluster --no-wait --chunk-jobs 1000
```

3. Example Compare MISO script in HiperGator:

```
python /home/eric.t.wang/Tools/pythonmodules/lib/python2.7/site-
packages/misopy/compare_miso.py \
--compare-samples /ufrc/berglund/npellit3/actD/data2/wt/
/ufrc/berglund/npellit3/2016_mouseRNAseq/data2/controls/ \
/ufrc/berglund/npellit3/2016_mouseRNAseq/data2/compare/
```

4. Example Sashimi Plot script in HiperGator:

```
python /home/eric.t.wang/Tools/pythonmodules/lib/python2.7/site-
packages/misopy/sashimi_plot/sashimi_plot.py \
--plot-event "chr8:84721643:84721818:-@chr8:84716141:84716288:-
@chr8:84713766:84713857:-"
/ufrc/berglund/npellit3/2016_mouseRNAseq/miso/annotations/SE_indexed \
/ufrc/berglund/npellit3/2016_mouseRNAseq/miso/sashimi_plot_settings.txt \
--output-dir /ufrc/berglund/npellit3/2016_mouseRNAseq/miso2/sashimi/
```

Appendix 4. DESeq2 Script in R Studio

```
library("DESeq2")
directory <- "/Users/nataliepellitier/Desktop/Thesis!!!/HTSeq/"
sampleFiles <- grep("furamidine", list.files(directory),value = TRUE)
sampleFilescont <- grep("saline", list.files(directory),value = TRUE)
sampleFiles <- c(sampleFiles,sampleFilescont)
sampleCondition <- sub("\\d.genecount","\\1",sampleFiles)
sampleTable <- data.frame(sampleName = sampleFiles, fileName = sampleFiles, condition =
sampleCondition)
ddsHTSeq <- DESeqDataSetFromHTSeqCount(sampleTable = sampleTable, directory =
directory, design = ~ condition)
ddsHTSeq$condition <- factor(ddsHTSeq$condition, levels=c("saline","furamidine"))
dds <- DESeq(ddsHTSeq)
res <- results(dds)
res <- res[order(res$padj),]
head(res)
summary(res)
plotMA(res, alpha=0.01, ylim=c(-5,5))
```

BIBLIOGRAPHY

Aldehag, Anna Sandin, Hans Jonsson, and Tor Ansved. "Effects of a hand training programme in five patients with myotonic dystrophy type 1." *Occupational Therapy International* 12.1 (2005): 14-27. Web.

Baldanzi, Sigrid, Francesca Bevilacqua, Rita Lorio, Leda Volpi, Costanza Simoncini, Antonio Petrucci, Mirco Cosottini, Gabriele Massimetti, Gloria Tognoni, Giulia Ricci, Corrado Angelini, and Gabriele Siciliano. "Disease awareness in myotonic dystrophy type 1: an observational cross-sectional study." *Orphanet Journal of Rare Diseases* 11.1 (2016): n. pag. Web.

Bellini, Massimo. "Gastrointestinal manifestations in myotonic muscular dystrophy." *World Journal of Gastroenterology* 12.12 (2006): 1821. Web.

Bird, Thomas D. "Myotonic Dystrophy Type 1." *GeneReviews [Internet]* (2015) Web.

Farahat, Abdelbasset A., Ekaterina Paliakov, Arvind Kumar, Alaa-Eldin M. Barghash, Fatma E. Goda, Hassan M. Eisa, Tanja Wenzler, Reto Brun, Yang Liu, W. David Wilson, and David W. Boykin. "Exploration of larger central ring linkers in furamidine analogues: Synthesis and evaluation of their DNA binding, antiparasitic and fluorescence properties." *Bioorganic & Medicinal Chemistry*19.7 (2011): 2156-167. Web.

Coonrod, Leslie A., Masayuki Nakamori, Wenli Wang, Samuel Carrell, Cameron L. Hilton, Micah J. Bodner, Ruth B. Siboni, Aaron G. Docter, Michael M. Haley, Charles A. Thornton, and J. Andrew Berglund. "Reducing Levels of Toxic RNA with Small Molecules." *ACS Chemical Biology* 8.11 (2013): 2528-537. Web.

Fardaei, M. "In vivo co-localisation of MBNL protein with DMPK expanded-repeat transcripts." *Nucleic Acids Research* 29.13 (2001): 2766-771. Web.

Foff, Erin Pennock, and Mani S. Mahadevan. "Therapeutics development in myotonic dystrophy type 1." *Muscle & Nerve* 44.2 (2011): 160-69. Web.

Gomes-Pereira, Mário, Thomas A. Cooper, and Geneviève Gourdon. "Myotonic dystrophy mouse models: towards rational therapy development." *Trends in Molecular Medicine* 17.9 (2011): 506-17. Web.

Lee, Johanna E., and Thomas A. Cooper. "Pathogenic mechanisms of myotonic dystrophy." *Biochemical Society Transactions* 37.6 (2009): 1281-286. Web

Mankodi, A. "Myotonic Dystrophy in Transgenic Mice Expressing an Expanded CUG Repeat." *Science* 289.5485 (2000): 1769-772. Web.

Mankodi, A. "Nuclear RNA Foci in the Heart in Myotonic Dystrophy." *Circulation Research* 97.11 (2005): 1152-155. Web.

Miller, J. W. "Recruitment of human muscleblind proteins to (CUG)n expansions associated with myotonic dystrophy." *The EMBO Journal* 19.17 (2000): 4439-448. Web.

Nakamori, Masayuki, et. al. "Oral administration of erythromycin decreases RNA toxicity in myotonic dystrophy." *Annals of Clinical and Translational Neurology*(2015): n. pag. Web.

Nakamori, Masayuki, et. al. "Neuromuscular disease: RNA splicing biomarkers identified for myotonic dystrophy." *Nature Reviews Neurology* 9.10 (2013): 543. Web.

Siboni, Ruth B., Micah J. Bodner, Muhammad M. Khalifa, Aaron G. Docter, Jessica Y. Choi, Masayuki Nakamori, Michael M. Haley, and J. Andrew Berglund. "Biological Efficacy and Toxicity of Diamidines in Myotonic Dystrophy Type 1 Models." *Journal of Medicinal Chemistry* 58.15 (2015): 5770-780. Web.

Sommerville, R. Brian, Margherita Guzzi Vincenti, Kathleen Winborn, Anne Casey, Nathan O. Stitziel, Anne M. Connolly, and Douglas L. Mann. "Diagnosis and management of adult hereditary cardio-neuromuscular disorders: A model for the multidisciplinary care of complex genetic disorders." *Trends in Cardiovascular Medicine* 27.1 (2017): 51-58. Web.

Thornton, Charles A., Eric Wang, and Ellie M. Carrell. "Myotonic dystrophy: approach to therapy." *Current Opinion in Genetics & Development* 44 (2017): 135-40. *Elsevier*. Web.

Van Agtmaal, Ellen L., Laurène M. André, Marieke Willemse, Sarah A. Cumming, Ingeborg D.g. Van Kessel, Walther J.a.a. Van Den Broek, Geneviève Gourdon, Denis Furling, Vincent Mouly, Darren G. Monckton, Derick G. Wansink, and Bé Wieringa. "CRISPR/Cas9-Induced (CTG·CAG)n Repeat Instability in the Myotonic Dystrophy Type 1 Locus: Implications for Therapeutic Genome Editing." *Molecular Therapy* 25.1 (2017): 24-43. Web.

Warf MB, Berglund JA: “MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T” RNA 2007, 13:2238-2251

Warf, M. B., M. Nakamori, C. M. Matthys, C. A. Thornton, and J. A. Berglund. "Pentamidine reverses the splicing defects associated with myotonic dystrophy." *Proceedings of the National Academy of Sciences* 106.44 (2009): 18551-8556. Web.

Winblad, Stefan, Christer Jensen, Jan-Eric Månsson, Lena Samuelsson, and Christopher Lindberg. "Depression in Myotonic Dystrophy type 1: clinical and neuronal correlates." *Behavioral and Brain Functions* 6.1 (2010): 25. Web.

Xia, Guangbin, Yuanzheng Gao, Shouguang Jin, S.h. Subramony, Naohiro Terada, Laura P.w. Ranum, Maurice S. Swanson, and Tetsuo Ashizawa. "Genome Modification Leads to Phenotype Reversal in Human Myotonic Dystrophy Type 1 Induced Pluripotent Stem Cell-Derived Neural Stem Cells." *Stem Cells* 33.6 (2015): 1829-838. Web.

Yum, Kevin, Eric T. Wang, and Auinash Kalsotra. "Myotonic dystrophy: disease repeat range, penetrance, age of onset, and relationship between repeat size and phenotypes." *Current Opinion in Genetics & Development* 44 (2017): 30-37. Web.

Methods Documentation

Bioo Scientific Molecular Adapters Documentation can be accessed at: http://www.biooscientific.com/Next-Gen-Sequencing/Illumina-RNA-Seq-Library-Prep-Kits/NEXTflex-qRNA-Seq-Kit

Love MI, Huber W and Anders S (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biology*, **15**, pp. 550. doi: 10.1186/s13059-014-0550-8.

MISO documentation can be accessed at: https://miso.readthedocs.org/en/fastmiso/

Shen S., Park JW., Lu ZX., Lin L., Henry MD., Wu YN., Zhou Q., Xing Y. rMATS: Robust and Flexible Detection of Differential Alternative Splicing from Replicate RNA-Seq Data. *PNAS*, 111(51):E5593-601. doi: 10.1073/pnas.1419161111

Simon Anders, Paul Theodor Pyl, Wolfgang Huber *HTSeq — A Python framework to work with high-throughput sequencing data*, Bioinformatics (2014), in print, online at doi:10.1093/bioinformatics/btu638

The technical data sheet for our RNA-sequencing kit can be accessed at: https://www.kapabiosystems.com/product-applications/products/next-generation-sequencing-2/rna-library-preparation-2/kapa-stranded-rna-seq-riboerase/