Identification of Key Long Non-Coding RNAs (LncRNAs) Effecting Alternative Splicing and Gene Expression in the Frontal Cortex of Myotonic Dystrophy Type 1 Patients

Tehreem Fatima

Shaker High School

445 Watervliet Shaker Rd, Latham, NY 12110

Acknowledgments

The successful completion of this project owes much to the invaluable guidance and unwavering support provided by my mentors at the Berglund Lab, within the University at Albany's RNA Institute. Dr. Berglund, Claudia Lennon, Cécilia Légaré, and Robert Merritt played vital roles in facilitating my research journey and equipped me with the essential foundational skills necessary for the setup and execution of this project. Additionally, I extend my heartfelt gratitude to my science research teacher at Shaker High School, Mr. Covert, whose mentorship and assistance were instrumental in shaping the scientific paper. Lastly, I wish to express my deep appreciation to my parents, Daud Javid and Shazia Daud, who, despite their lack of interest in research, offered their unwavering support and encouragement, enabling me to pursue my academic endeavors.

Abstract

Myotonic dystrophy type 1 (DM1) is a complex genetic multisystemic disorder marked by the expansion of CTG repeats within the 3' untranslated region of the dystrophia myotonic protein kinase (DMPK) gene. DM1 is associated with a severe neuromuscular phenotype, encompassing early-onset ataxia, dysarthria, muscle weakness, and exercise intolerance. The sophisticated nature of DM1 necessitates a nuanced exploration of its molecular groundworks. In this study, I embarked on an investigation into the potential roles of long non-coding RNAs (lncRNAs) in DM1 pathophysiology. LncRNAs, once regarded as genomic "junk," have emerged as critical regulators of gene expression and alternative splicing, making them intriguing candidates for understanding DM1's underlying mechanisms. Within the realm of alternative splicing in DM1, I identified NUTM2A-AS1 as a potential player, substantially influencing splicing patterns in DM1-affected brains. Importantly, this lncRNA exhibited non-coding repeat expansions, suggesting a central role in the development of DM1.

Additionally, I unveiled other lncRNAs with their associated genes and proteins, such as KHDRBS3 and HDAC2, each offering unique insights into the complex world of alternative splicing. Regarding differential gene expression in DM1, I pinpointed several lncRNAs and their corresponding genes or proteins, including MAP6, FOSL2, and HLA-DQB1. These discoveries illuminate the intricate interplay of genes and proteins, shedding light on the multifaceted nature of DM1.

Table of Contents

Page #

List of Figures

Page #

Introduction

Myotonic dystrophy type 1 (DM1) stands out as the most prevalent muscular dystrophy in adults, marked by its status as a genetic multisystemic disorder (Serra et al., 2014). This condition results from the expansion of CTG repeats within the 3' untranslated region (UTR) of the dystrophia myotonic protein kinase (DMPK) gene. Its primary presentation concerns myotonic muscular disorder, characterized by the gradual onset of muscle weakness and atrophy due to myotonia. Importantly, DM1 encompasses a spectrum of anomalies that affect various organ systems, notably influencing the central nervous system and the brain (Brook et al., 1992).

Within the realm of the brain, DM1 can induce cognitive disturbances closely associated with atypical brain activity patterns. This includes phenomena such as mental fatigue, daytime sleepiness, forgetfulness, and confusion. Intriguingly, individuals afflicted with DM1 display nerve cell counts that approximate those of the general population. However, the underlying genetic mutation linked to DM1 instigates molecular and functional anomalies in the nerves of DM1 patients (Cumming et al., 2020). This amalgamation of genetic, physiological, and clinical intricacies underpins the complexity of DM1, necessitating a comprehensive approach to understanding and addressing its effects on individuals.

The precise molecular mechanisms underpinning the neurobiological alterations observed in DM1 within the brain continue to be a subject of investigation and scientific curiosity (Otero et al., 2021). In the realm of scientific exploration, notable strides were made in the understanding of specific proteins, such as Muscleblind proteins (MBNL), which have unveiled valuable insights into the potential replication of neuropathological defects. Significantly, the *in*

vivo loss of function of RNA-binding MBNL proteins has been associated with developing the distinctive muscular abnormalities characteristic of DM1 (Tahraoui-Bories et al., 2023).

Nonetheless, a comprehensive grasp of the biological mechanisms at play remains elusive. The vast transcriptome of the genome presents a perplexing challenge. If we conservatively estimate that only 10% of the genome is subject to selective pressure, while about 80% of it is transcribed, it suggests that at least 87% of the transcribed regions may be deemed as non-functional. (Palazzo et al., 2020). The multifaceted puzzle of genome functionality and the extent of "junk" RNA within it continue to intrigue scientists, adding a layer of complexity to the understanding of the molecular landscape in DM1. In recent years, an expanding body of evidence has cast a revealing light on the biological importance of long non-coding RNAs (lncRNAs) across various disorders, including muscular dystrophies (Erriquez et al., 2013). This evolution in scientific understanding challenges the previous notion of lncRNAs as "junk" RNA, long believed to lack function. It is now evident that functional lncRNAs not only exist but also proliferate within the intricate human biological landscape (Palazzo et al., 2020).

LncRNAs represent a captivating category of RNA molecules, distinguished by their length, exceeding 200 nucleotides, and remarkable diversity. Their classification hinges on variables such as genomic location and orientation, embracing sense, antisense, bidirectional, intronic, or intergenic lncRNAs. Additionally, although they are non-coding RNAs, lncRNAs hold a multitude of functions within the human body, including regulation during development and involvement in almost all levels of gene expression and cellular functions, including chromosomal dosage compensation, chromatin modification, cell cycle regulation, control of

imprinting, alternative splicing, intracellular trafficking, cellular differentiation, and reprogramming of stem cells (Erriquez et al., 2013) (Figure 1.).

Furthermore, it is worth highlighting that lncRNAs exhibit significant conservation in both their secondary and tertiary structures. A multitude of recent studies have consistently emphasized the strong association between lncRNAs and the initiation, progression, diagnosis, and therapeutic strategies for various diseases (Zhang et al., 2018).

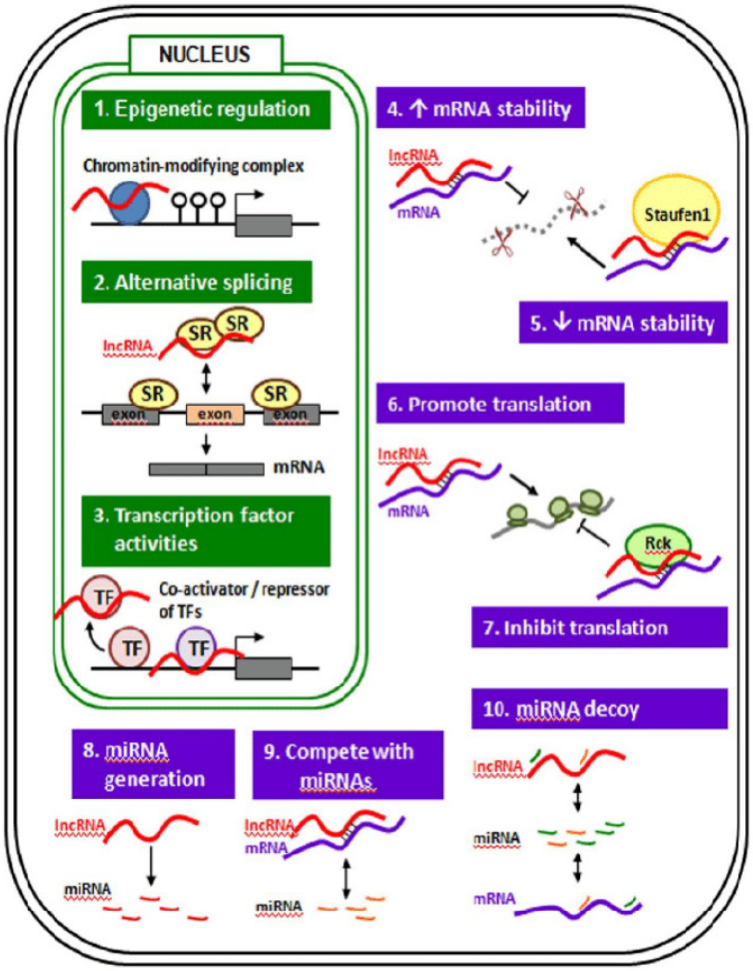

Fig 1. Gene regulation mechanisms by lncRNAs and various events such as alternative Splicing (2) and transcript expression (3) with the nucleus of Eukaryotic cells (Li et al., 2017)

As previously alluded, lncRNAs possess functional relevance in alternative splicing and gene expression. It is crucial to acknowledge that DM1 is characterized by the dysregulation of alternative splicing and gene expression (Nakamori et al., 2013). While the regulatory role of lncRNAs has been delineated in the context of muscular dystrophy muscle (Koscianska et al., 2021), their prevalence and specific functions within DM1 brains have remained unexplored. However, it is noteworthy that lncRNAs have been implicated in several neurodegenerative diseases, such as Huntington's and Alzheimer's (Koscianska et al., 2021). Given that DM1 is a neuromuscular disorder, the involvement of lncRNAs in the modulation of splicing and gene expression within the disease's context is a promising avenue of investigation.

Hypothesis

The preliminary introduction raises the following primary hypothesis:

1. LncRNAs will significantly affect alternative splicing and gene expression within the frontal cortex of DM1 patients.

Materials List

- FASTQC Version 0.12.0
- University at Albany’s HPCC (High Performance Computing Cluster)
- rMATS (Multivariate Analysis of Transcript Splicing): Version 4.1.2
- STAR (Spliced Transcripts Alignment to a Reference): Version 2.7.11a

- DESeq2 1.42.0 (Differential gene expression analysis based on the negative binomial distribution)
- R version 4.2.2
- Rstudio 2023.06.1 Build 524
    - AnnotationDbi 1.64.0 (Manipulation of SQLite-based annotations in Bioconductor)
    - Org.Hs.eg.db (Genome wide annotation for Human)
    - EnhancedVolcano 1.20.0 (Publication-ready volcano plots with enhanced coloring and labeling)
    - Pheatmap 1.0.12 (Pretty Heatmaps)
    - Maser (Mapping Alternative Splicing Events to pRoteins) (3.18)
    - Dplyr 1.1.3 (A Grammar of Data Manipulation)
- Ensemble Biomart (Ensembl release 110 - July 2023)
- Gene Cards (September 2023)
- UCSC Genome Browser (September 2023)
- LNCipedia version 5.2

Methods

All data was acquired from Otero et. al (2021) and they acquired human autopsy samples from the frontal cortex (Brodmann area 10) of post-mortem individuals with DM1, procured from the following esteemed institutions, Stanford University, The Research Resource Network Japan, University of Rochester Medical Center, and the NIH Biobank. These tissue specimens were subjected to thorough analysis and quantification by Otero et al. (2021) for a research

project investigating transcriptome alterations in myotonic dystrophy frontal cortex. All acquired RNA data samples exhibited an RQN score of less than 4, underscoring the high quality and fidelity of the dataset.

The dataset consisted of 21 DM1 patients, comprising 12 females and 9 males. For comparative analysis, an additional group of 8 control patients, including 4 males and 4 females, was included. The ages of the individuals in the dataset spanned from 39 to 83 years, with an average age of 64 years, demonstrating the diversity and range of the studied population.

The data was meticulously formatted into FASTQ files and met a rigorous quality control assessment via FASTQC Version 0.12.0. I downloaded this data from the *GEO* (Gene Expression Omnibus)/*SRA* (Sequence Read Archive) of the published article and using the data, I employed a series of analytical methodologies. All FASTQ file reads were mapped to the hg19 genome using STAR (Spliced Transcripts Alignment to a Reference) Version 2.7.11a. All preliminary analyses were run using the University at Albany's High Performance Computing Cluster (HPCC). All RNA-Seq libraries used were sequenced to a depth of at least 88 million reads which provided sufficient coverage for analyses of gene expression and alternative splicing. The procedures for alternative splicing analysis and the evaluation of differential expression are delineated in Figure 2, providing a comprehensive exposition of the analytical workflow.

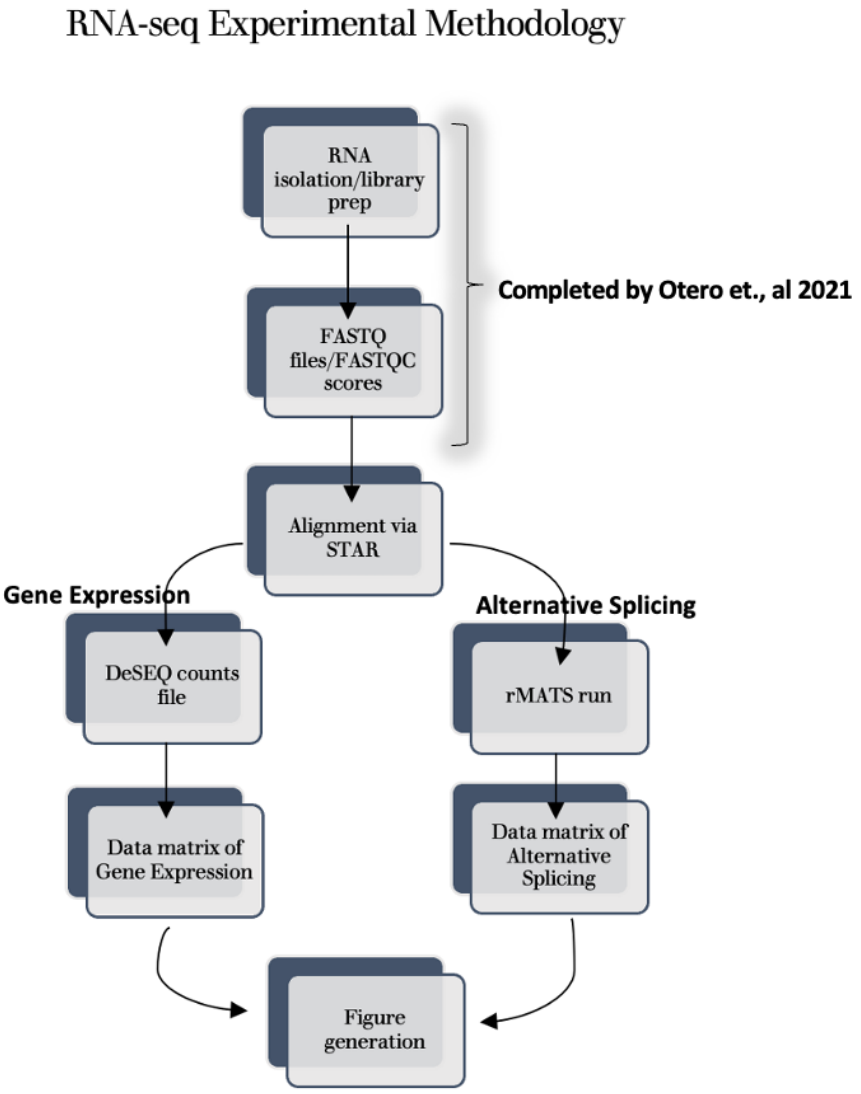

Fig 2. The initial data analysis and filtering process followed a structured pipeline, with specific software versions used for each step (please refer to the "Methods" section for detailed program version information).

Before proceeding to the generation of figures, supplementary analysis was conducted on the data matrices encompassing both gene expression and alternative splicing. A pivotal step involved the stringent filtration of data, wherein *significance* was defined as a $\Delta PSI \geq |0.1|$ and a

rank-sum p-value ≤ 0.01 for splicing events. Additionally, for gene expression, significance was determined by a log_2 fold change greater than |2| and a p-value ≤ 0.01.

Alternative Splicing

Upon the completion of the splicing data analysis, the results were exported into five distinct filtered CSV files, categorizing the splicing events into A3SS (alternative 3' splice site), A5SS (alternative 5' splice site), RI (intron retention), MXE (mutually exclusive exon), and SE (skipped exon) events. Subsequently, all anomalies derived from these five splicing event categories were combined into a single Excel sheet. Within this consolidated data, Gene IDs that remained unidentified were extracted. To rectify this, the Ensemble IDs associated with these unidentified entities were employed to perform a detailed characterization utilizing Ensemble Biomart (Ensembl release 110 - July 2023). This comprehensive classification encompassed identifying these 158 entities as genes, pseudogenes, novel transcripts, lncRNAs, or a group of transcripts that remained unclassified. Furthermore, the transcripts that were initially unidentified were subjected to an additional identification process through Gene Cards (September 2023). Finally, 78 significant lncRNAs were extracted and organized into a separate Excel sheet for further analysis and study.

Differential Gene Expression (DGE)

Following the completion of the analysis of the DGE (Differential Gene Expression) data, the resulting information was organized into a refined CSV file. Subsequently, all irregularities and Gene IDs that remained unclassified were extracted from this data. These unidentified entities were then scrutinized using their Ensemble IDs for in-depth

characterization, with reference to Ensemble Biomart (Ensembl release 110 - July 2023). The characterization encompassed classifying these 100 anomalies into genes, pseudogenes, novel transcripts, lncRNAs, or a group of transcripts that defied classification.

For the transcripts that persisted as unidentified after this initial step, an additional identification process was carried out using Gene Cards (September 2023). Finally, 51 significant lncRNAs were identified and cataloged into a separate Excel sheet for further research and analysis.

Figure Generation

Initial data analysis encompassed the generation of figures for both splicing and gene expression datasets. This rigorous analytical process was executed using R version 4.2.2, and the coding was performed in Rstudio (Version 2023.06.1 Build 524). The visualization outputs included Principal Component Analysis (PCA) plots, Volcano plots, and Heatmaps, providing a comprehensive view of the entire dataset. These figures effectively captured and conveyed fundamental patterns and variances within the data.

Within the realm of alternative splicing analysis, violin plots and modified sashimi plots were attentively crafted, explicitly focusing on all 158 significant long non-coding RNAs (lncRNAs) identified. These visualizations offered insights into the splicing patterns and characteristics associated with these lncRNAs. Furthermore, in the context of Differential Gene Expression (DGE) analysis, Normalized Count Plots were constructed, spotlighting all 100 significant lncRNAs. These plots facilitated the visualization of the expression levels and variability of these lncRNAs across the studied samples, contributing to a deeper understanding

of their differential expression patterns. Subsequently, based on a previously defined significance, the pool of lncRNAs was judiciously narrowed down to 78 for alternative splicing and 51 for gene expression, ensuring a focused and insightful analysis of the most pertinent lncRNAs.

Review of Literature

In a meticulous effort to enhance the understanding of the 78 lncRNAs significantly impacting splicing and the 51 lncRNAs influencing gene expression, a comprehensive literature search was diligently conducted. The primary goal was to identify any prior research articles that shed light on the significance of these lncRNAs in the context of DM1 brains. The literature search yielded valuable findings.

For the 78 lncRNAs influencing alternative splicing in DM1 brains, 1 specific lncRNA was discovered to have been previously identified by Hiroyuki Ishiura et al. (2019) as playing a significant role in this context. Additionally, the investigation revealed 4 proteins/genes closely associated with this particular lncRNA, all of which are known to have a substantial impact on DM1 brains through their interaction with the lncRNA. These lncRNAs included *Lnc-PAPOLA-1*, *Lnc-SLC25A32-3*, *Lnc-KHDRBS3-8*, *HDAC2-AS2* and the 4 genes/proteins included PAPOLA (Batra et al., 2014), SLC25A32 (Marra et al., 2021), KHDRBS3 (Feng et al., 2021), HDAC2 (Grande et al., 2021) which were identified to play a role in splicing within DM1 brains.

Conversely, regarding the 51 lncRNAs influencing gene expression, no specific lncRNAs were found to be directly linked to previous research. However, the inquiry did uncover 3

proteins/genes for which the identified lncRNAs played a role in modulating gene expression within DM1 brains. These lncRNAs included *Lnc-MAP6-2, FOSL2-AS1, HLA-DQB1-AS1, a*nd the 3 genes/proteins included MAP6 (Velázquez-Bernardino et al., 2011), FOSL2 (Batra et al., 2014), HLA-DQB1 (Martínez-Rodríguez et al. 2003).

It is noteworthy that many of the identified long non-coding RNAs (lncRNAs) and the proteins/genes closely related to them have been found to be prevalent in a multitude of other diseases, including Alzheimer's, Huntington's, and various forms of cancer (Johnson, 2012). This cross-disease association underscores the complex and multifaceted roles these lncRNAs and proteins/genes play in the context of various neurological and oncological conditions, offering potential insights into common molecular mechanisms and therapeutic targets shared among these diseases.

Gene Ontology

The Ensemble IDs derived from the analysis of the 78 significant long non-coding RNAs (lncRNAs) related to splicing, as well as the 51 significant lncRNAs pertaining to gene expression, were independently inputted into both PANTHER18.0 and MetaScape (release 2023-09-01). This approach aimed to conduct a comprehensive gene ontology analysis to gain deeper insights into the biological functions, pathways, and molecular processes associated with these lncRNAs in the context of the studied dataset.

Results

Alternative Splicing

In the context of alternative splicing in DM1 brains, one particular lncRNA stood out with significance: NUTM2A-AS1 (NUTM2A antisense RNA 1). This lncRNA is notable in the alternative splicing processes associated with DM1.

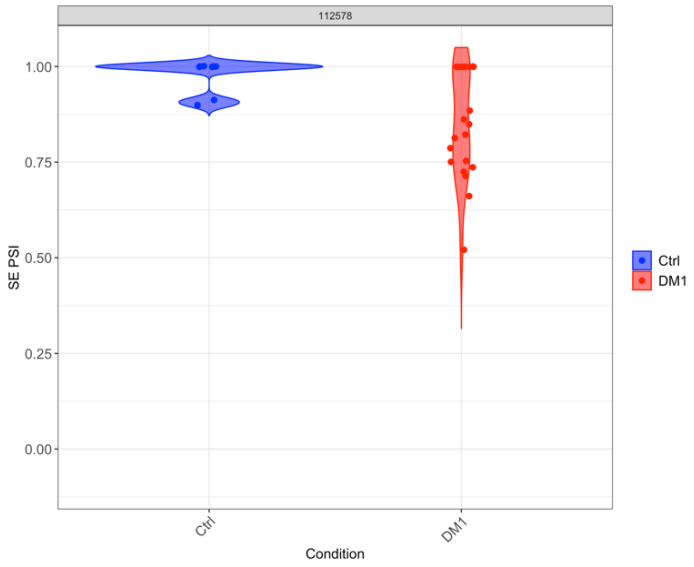

Fig 3: Violin Plot representing splicing in the skipped exon event between DM1 and control in NUTM2A-AS1. Controls are seen clustered towards 1.00 PSI while, DM1 range between 0.50 PSI -1.00 PSI.

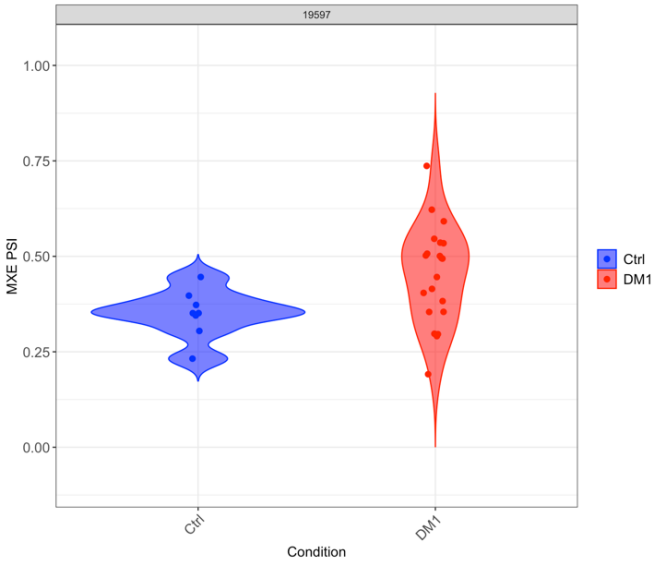

Fig 4: Violin Plot representing splicing in the mutually exclusive exon event between DM1 and control in NUTM2A-AS. Controls are seen clustered between 0.25 PSI - 0.50 PSI while DM1 range between 0.25 PSI -0.75 PSI.

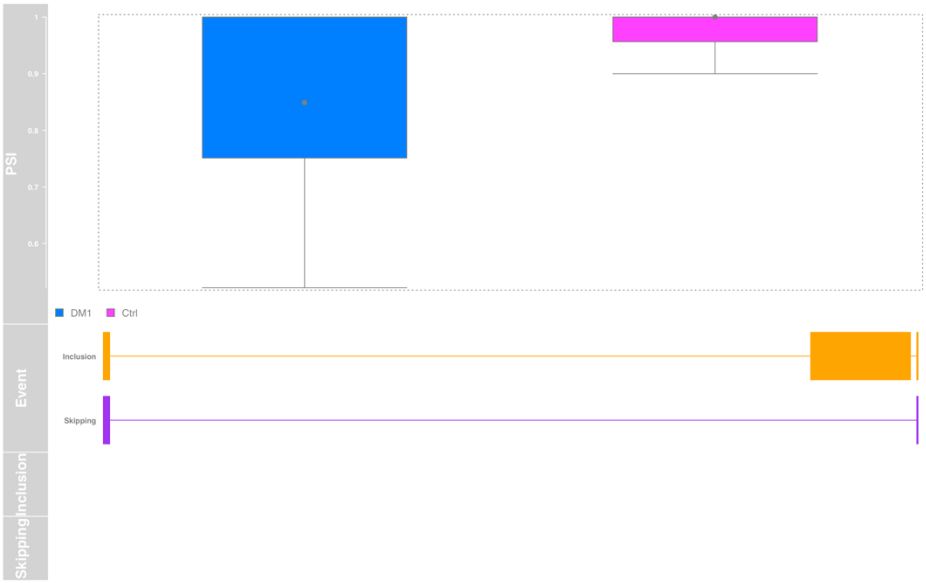

Fig 5: Modified Sashimi Plot representing splicing in the skipped exon event between DM1 and control in NUTM2A-AS1. This plot is a box plot quantification of Fig. 3. Controls are seen ranging between 0.95 PSI - 1.00 PSI while DM1 range between 0.75 PSI - 1.00 PSI.

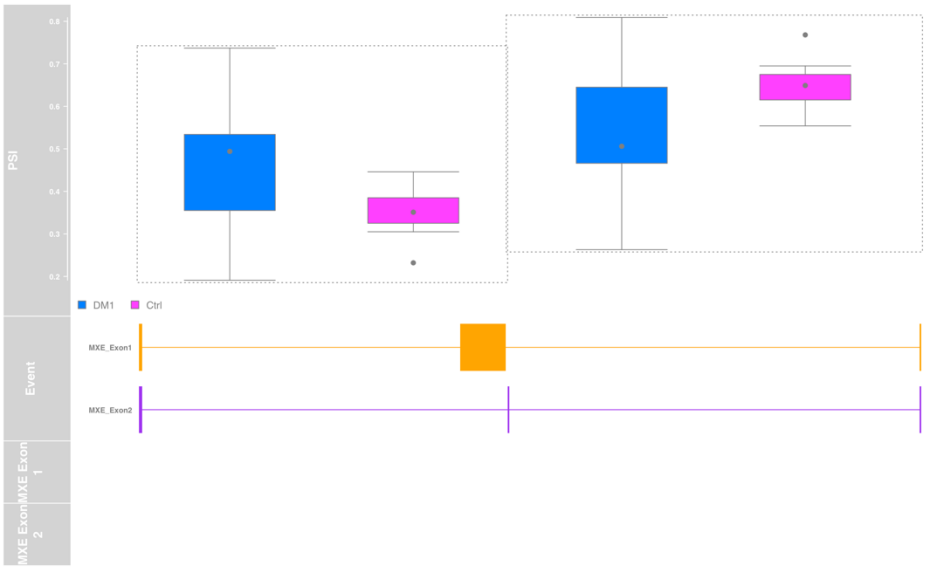

Fig 6: Modified Sashimi Plot representing splicing in the mutually exclusive exon event between DM1 and control in NUTM2A-AS1 This plot is a box plot quantification of Fig. 4. Within the mutually exclusive exon 1 controls are seen ranging between 0.30 PSI - 0.375 PSI while DM1 range between 0.35 PSI - 0.55 PSI. While, within the mutually exclusive exon 1 controls are seen ranging between 0.625 PSI - 0.675 PSI while DM1 range between 0.45 PSI - 0.65 PSI.

The significance of NUTM2A-AS1 is evident within Figures 1-6, as these figures illustrate apparent differences in splicing patterns between individuals with DM1 and control subjects. These differences emphasize the pivotal role of NUTM2A-AS1 in mediating the distinct splicing patterns observed in the context of DM1. Moreover, a study conducted by Hiroyuki Ishiura and colleagues (Hiroyuki Ishiura et al., 2019) focusing on non-coding repeat expansions causing various neuromuscular diseases, including myotonic dystrophies, sheds further light on the importance of NUTM2A-AS1. The research findings revealed the presence of significant non-coding CGG repeat expansions within NUTM2A-AS1. The research reveals significant non-coding CGG repeat expansions within NUTM2A-AS1, known to be associated with neuromuscular disease pathogenesis.

Despite DM1 having CTG repeats, the *significance* of NUTM2A-AS1's splicing differences in DM1 individuals compared to controls is evident in the figures above. Furthermore, the study uncovers that another lncRNA within the NUTM2 family, NUTM2B-AS1, has CGG repeats in 1,000 control subjects, ranging from 3 to 16. This highlights the potential role in regulating NUTM2A-AS1 and its' relevance to the development of DM1. These findings may provide valuable insights into the genetic and molecular underpinnings of DM1 and suggest that NUTM2A-AS1, along with related non-coding repeat expansions, plays a crucial role in the disease's etiology and pathophysiology. Further research in this area may help to unravel the intricate mechanisms by which NUTM2A-AS1 contributes to the disease and may offer potential avenues for therapeutic interventions.

Furthermore, the inquiry revealed several other lncRNAs and their corresponding genes or proteins, which are of relevance in the context of DM1:

Lnc-**PAPOLA**-1, Lnc-**SLC25A32**-3, Lnc-**KHDRBS3**-8, **HDAC2**-AS2

Significance of lncRNA proximal proteins/genes/enzymes:

1. **PAPOLA**: A gene that encodes for the poly(A) polymerase alpha (PAPα) enzyme, which is involved in the addition of the poly(A) tail to messenger RNA (mRNA) molecules. In the context of DM1, research has revealed that compound depletion of Mbnl1 and Mbnl2, which are RNA-binding proteins, leads to significant dysregulation of PAPOLA. The dysregulation is observed along the proximal-to-distal gradient within the brains affected by DM1 (Batra et al., 2014).

2. **SLC25A32:** A gene that encodes a mitochondrial transporter protein. In the context of DM1, it is noteworthy that neuromuscular disorders have also been associated with SLC25A32 deficiency. In a study by Marra et al. 2021, a novel homozygous mutation in SLC25A32 was identified in a patient with a particularly severe neuromuscular phenotype. This phenotype included early-onset ataxia, myoclonia (involuntary muscle jerks), dysarthria (difficulty in speech articulation), muscle weakness, and exercise intolerance.

3. **KHDRBS3:** A gene that encodes the RNA-binding protein KH domain-containing, RNA-binding, signal transduction-associated protein 3. In a study by Feng et al. in 2021, the complexity and graded regulation of neuronal cell-type-specific alternative splicing were

explored through single-cell RNA sequencing. Within this research, it was observed that Khdrbs3 predominantly represses exon inclusion. This observation is consistent with the preferential skipping of its target exons, particularly in glutamatergic neurons. Notably, another protein of relevance in the context of DM1, Mbnl2, exhibits preferential expression and higher activity in glutamatergic neurons compared to GABAergic neurons. Consequently, the splicing activation or repression mediated by these RNA-binding proteins (RBPs), such as KHDRBS3 and Mbnl2, results in a splicing profile specific to glutamatergic neurons.

4. **HDAC2:** Histone deacetylase 2 is an essential chromatin-modifying enzyme. HDAC2's primary function is to repress gene transcription by removing acetyl groups from histones. In the context of DM1 pathophysiology, HDAC inhibitors have been explored as potential therapeutic agents. These inhibitors have been shown to increase the expression of Muscleblind-like 1 (MBNL1), a protein deficient in DM1, and partially rescue aberrant splicing patterns in DM1-affected individuals. This suggests that HDAC2 may have a direct role in the pathophysiology of DM1 (Grande et al., 2021).

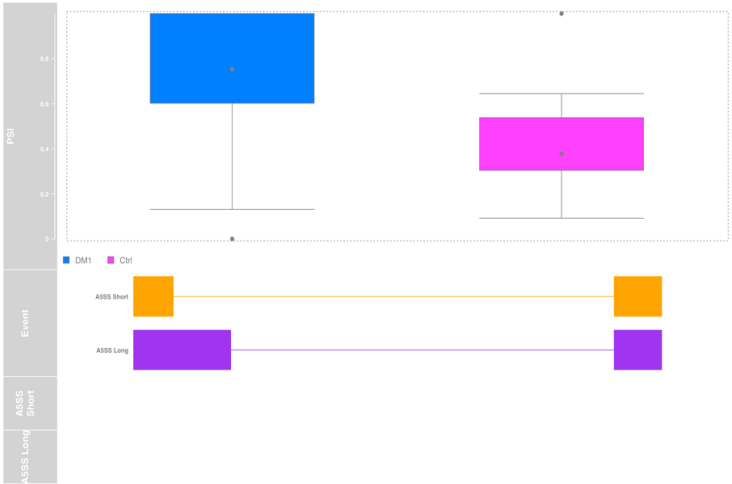

Fig 7: Modified Sashimi Plot representing splicing in the skipped exon event between DM1 and control in HDAC2-AS2. Controls are range between 0.25 PSI - 0.50 PSI while DM1 range between 0.60 PSI - 1.00 PSI.

Differential Gene Expression

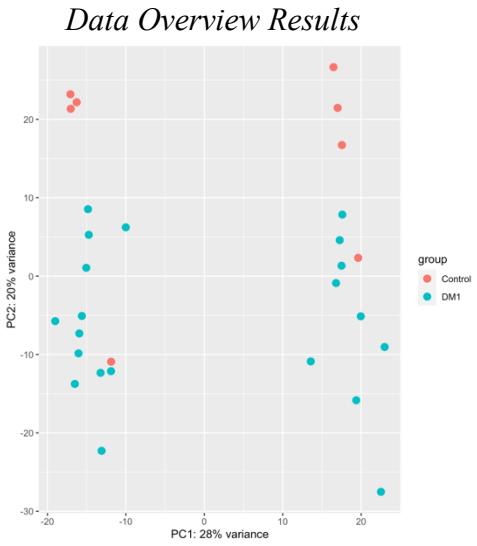

Fig 8: Principal Component Plot (PCA Plot) depicting the clustering between DM1 and control, and another clustering based on sex. On the right, 4 Ctrl males and 9 DM1 males are clustered. On the left, 4 Ctrl females and 12 DM1 females are clustered.

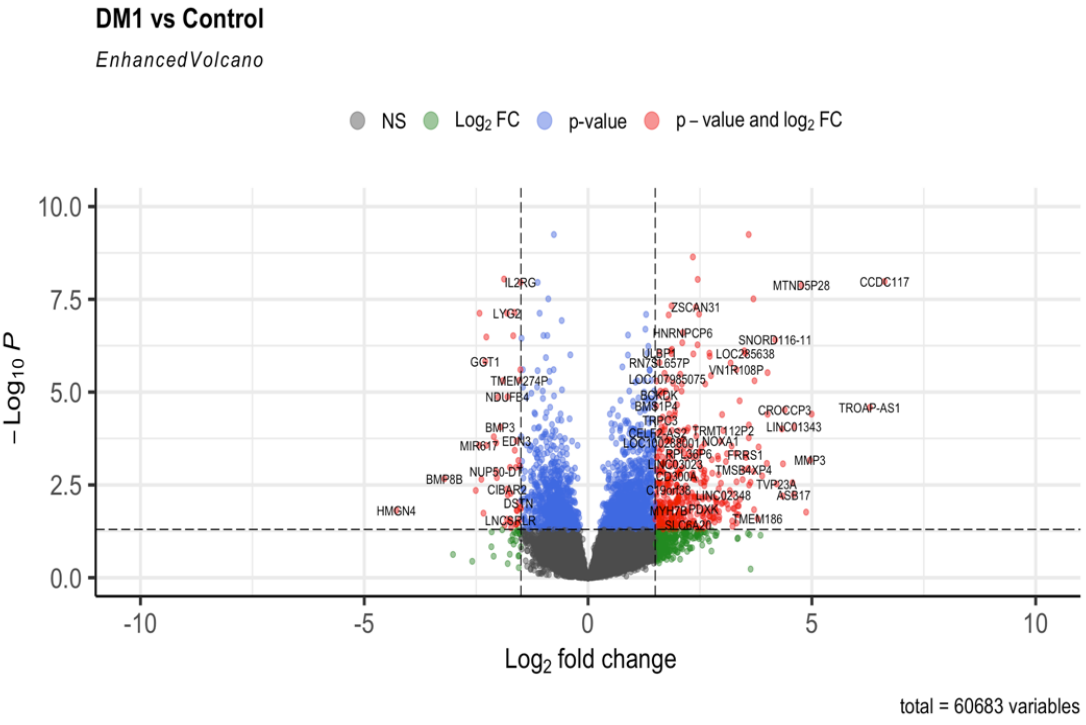

Fig 9: Volcano Plot depicting variables that meet cutoffs of significance. LncRNAs are seen labeled with genes and highlighted with yellow.

In the realm of gene expression in the context of DM1, the following long non-coding RNAs (lncRNAs) and their corresponding genes or proteins were identified:

Lnc-**MAP6**-2, **FOSL2**-AS1, **HLA-DQB1**-AS1

1. **MAP6:** Microtubule-associated Protein 6 (MAP6) is closely linked to microtubules, essential components of the cellular cytoskeleton. A study by Velázquez-Bernardino et al., 2011 aimed to understand the impact of DM1-associated CTG repeats on neuronal function. MAPs, including MAP6, are crucial in maintaining microtubule stability, which is vital for neuronal functions. The study suggests that CTG repeats may influence MAP expression, especially MAP6, and this alteration may contribute to neurite outgrowth deficits in CTG90 cells, a DM1 cellular model. Reduced MAP2 levels in the hippocampus of a DM1 mouse model also highlight the need to explore how CTG repeats affect MAP regulation, potentially relevant to DM1 pathophysiology (see Fig. 10 for ***Lnc-MAP6-2*** expression differentiation).

2. **FOSL2:** A gene that encodes a protein known as Fos-related antigen 2 (FRA-2) has significance in the context of DM1. Research has uncovered that the simultaneous reduction of Mbnl1 and Mbnl2, which are RNA-binding proteins, results in substantial dysregulation of Fosl2. This dysregulation manifests as a gradient shift from distal to proximal within the brains affected by DM1 (Batra et al., 2014). (***FOSL2-AS1*** and its differentiation in gene expression levels is depicted in Fig. 11)

3. **HLA-DQB1:** HLA-DQB1 encodes a subunit of the human leukocyte antigen (HLA) class II complex. A study by Martínez-Rodríguez et al. 2003 found reduced Hypocretin-1 (Orexin-A)

levels in the cerebrospinal fluid of patients with Myotonic Dystrophy and Excessive Daytime Sleepiness, which has implications for narcolepsy. Narcolepsy, characterized by excessive daytime sleepiness and REM sleep issues like cataplexy, likely results from hypocretin system dysfunction. More than 90% of narcoleptic patients with HLA-DQB1 positivity and cataplexy have undetectable Hypocretin-1 levels in their cerebrospinal fluid. Moreover, neuropathological investigations in narcoleptic brains have revealed either the absence or a significant reduction in the number of hypocretin neurons in the posterior hypothalamus. (see Fig. 12 for ***HLA-DQB1-AS1*** gene expression differentiation).

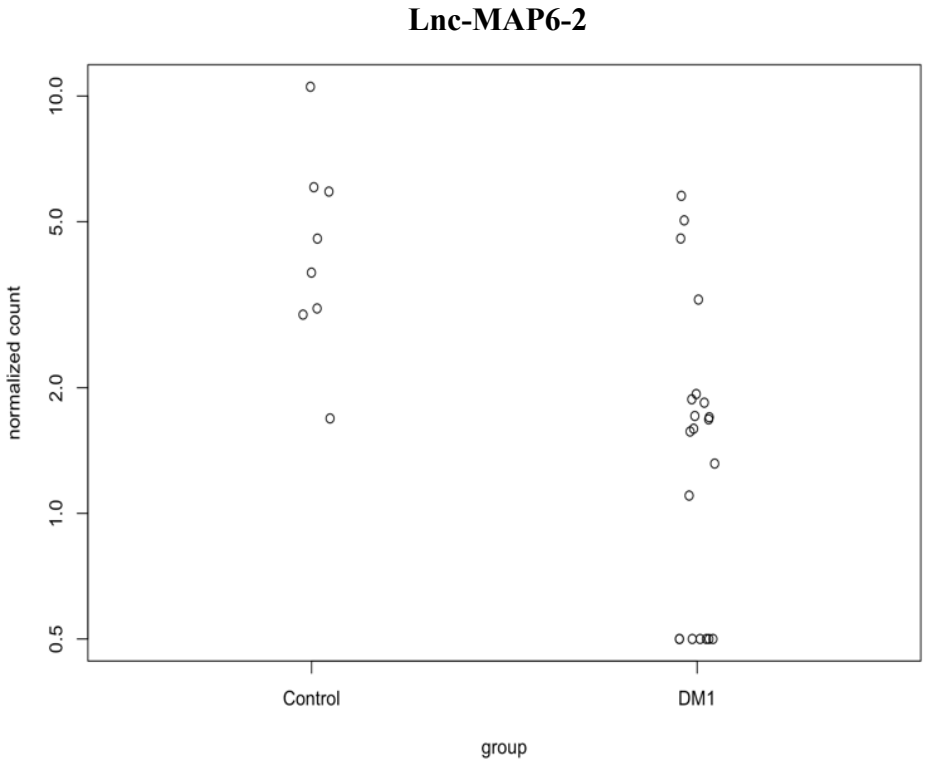

Fig 10: Normalized count plot for *Lnc-MAP6-2***.** Controls are seen to be higher in counts while DM1 depict low clustering at 0.5.

FOSL2-AS1

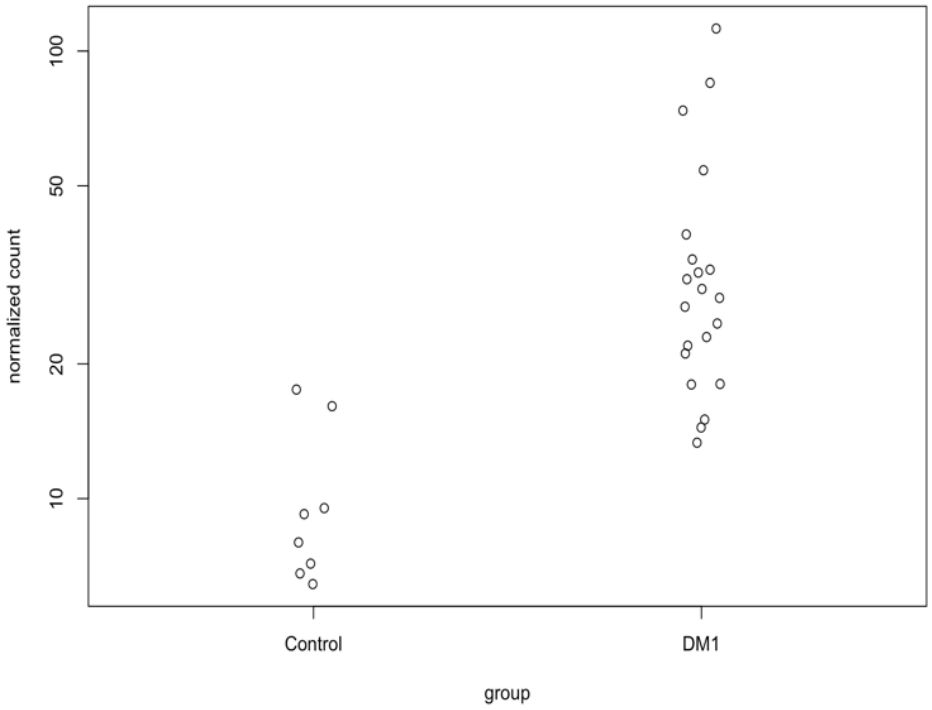

Fig 11: Normalized count plot for *FOSL2-AS1* is higher in DM1 than in controls.

HLA-DQB1-AS1

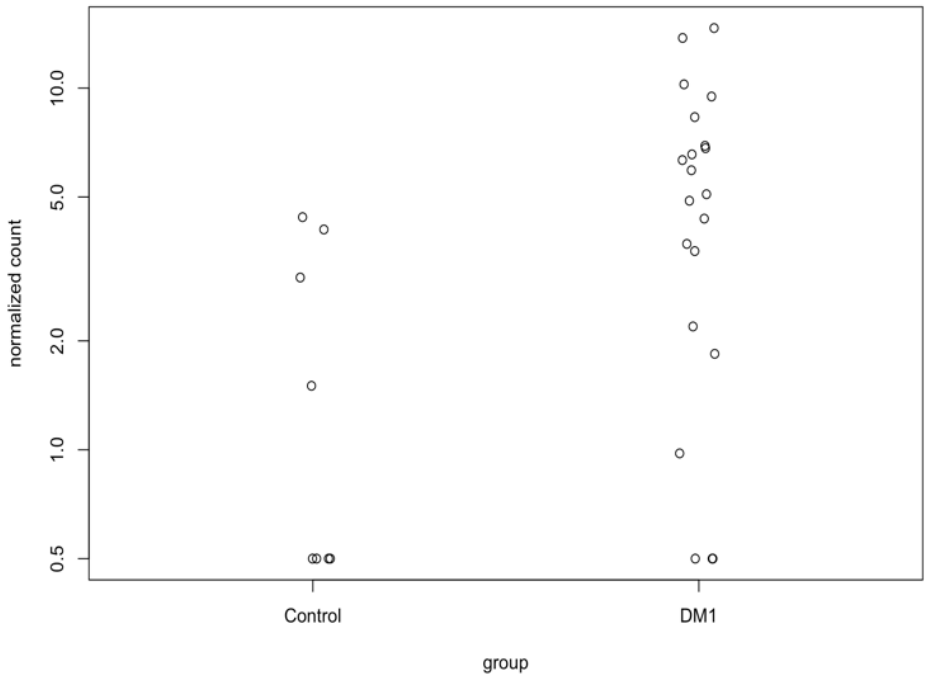

Fig 12: Normalized count plot for *HLA-DQB1-AS1* is higher in DM1 than in controls.

Discussion

It is essential to acknowledge that due to limitations in data availability and the relatively small number of identified lncRNAs, I was unable to generate results for Gene Ontology using PANTHER18.0 or MetaScape (release 2023-09-01). This limitation underscores the need for further comprehensive studies and data acquisition to gain a more profound understanding of the functional roles of these lncRNAs.

The broader context reveals that lncRNAs are not unique to Myotonic Dystrophy; they play significant roles in various disease contexts, including cancer and neurodegenerative conditions. Notably, in the case of breast cancer, lncRNAs have been associated with both the development and progression of the disease, suggesting their potential as valuable prognostic biomarkers. In a study involving samples from breast cancer patients, a distinctive five-lncRNA signature was identified. This signature was significantly linked to disease-free survival and effectively stratified patients into high- and low-risk groups. Remarkably, this predictive capability held regardless of subtype classification and adjuvant treatment, indicating the promise of these five lncRNAs as robust prognostic indicators for breast cancer (Li et al., 2018).

Additionally, lncRNAs have been recognized as influential players in neurodegenerative diseases such as Huntington's (HD). In a specific study by Johnson (2012), the potential significance of lncRNAs in HD was explored, leveraging existing microarray data to identify seven novel lncRNAs that exhibit dysregulation in HD brains. This finding underscores the expanding role of lncRNAs in the pathophysiological processes of neurodegenerative conditions, including well-known disorders like Alzheimer's disease (AD) and HD.

Conclusion

My research project was undertaken with the overarching objective of gaining deeper insights into the roles of long non-coding RNAs (lncRNAs) within the human biological landscape, specifically focusing on their impact on myotonic dystrophy type 1 (DM1) within the brain. Through this exploration, I successfully identified NUTM2A-AS1 as a lncRNA that appears to exert a significant influence on the alternative splicing processes associated with DM1. Furthermore, I pinpointed four additional lncRNAs affecting splicing and four influencing gene expression, each corresponding with specific genes known to reciprocally impact lncRNAs within the frontal cortex.

The identification of these lncRNAs holds promise for the scientific community, as it offers potential avenues for *in vivo* experimentation. These lncRNAs may serve as valuable biomarkers for DM1, helping clinicians to better understand and diagnose the disease. Additionally, the potential therapeutic implications of these lncRNAs provide a hopeful prospect for developing targeted treatments, thereby improving the quality of life for individuals affected by DM1.

References

Batra, R., Charizanis, K., Manchanda, M., Mohan, A., Li, M., Finn, Dustin J., … Swanson, Maurice S. (2014). Loss of MBNL Leads to Disruption of Developmentally Regulated Alternative Polyadenylation in RNA-Mediated Disease. *Molecular Cell*, *56*(2), 311–322. https://doi.org/10.1016/j.molcel.2014.08.027

Brook, J. David., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., … Juvonen, V. (1992). Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. *Cell*, *68*(4), 799–808. https://doi.org/10.1016/0092-8674(92)90154-5

Cumming, K. (2020, June 19). Myotonic Dystrophy and the Brain: Causes, Effects and Treatment. Retrieved from Myotonic Dystrophy Foundation website: https://www.myotonic.org/digital-academy/myotonic-dystrophy-and-brain-causes-effects-and-treatment#:~:text=DM%20can%20cause%20mental%20fatigue

Erriquez, D., Perini, G., & Ferlini, A. (2013). Non-Coding RNAs in Muscle Dystrophies. *International Journal of Molecular Sciences*, *14*(10), 19681–19704. https://doi.org/10.3390/ijms141019681

Feng, H., Moakley, D. F., Chen, S., McKenzie, M. G., Menon, V., & Zhang, C. (2021). Complexity and graded regulation of neuronal cell-type–specific alternative splicing revealed by single-cell RNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(10). https://doi.org/10.1073/pnas.2013056118

Grande, V., Hathazi, D., O'Connor, E., Marteau, T., Schara-Schmidt, U., Hentschel, A., … Roos, A. (2021). Dysregulation of GSK3β-Target Proteins in Skin Fibroblasts of Myotonic

Dystrophy Type 1 (DM1) Patients. *Journal of Neuromuscular Diseases*, *8*(4), 603–619. https://doi.org/10.3233/jnd-200558

Hiroyuki Ishiura, Shibata, S., Yoshimura, J., Suzuki, Y., Qu, W., Doi, K., … Ryo Ohtomo. (2019). Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease. *Nature Genetics*, *51*(8), 1222–1232. https://doi.org/10.1038/s41588-019-0458-z

Johnson, R. (2012). Long non-coding RNAs in Huntington's disease neurodegeneration. *Neurobiology of Disease*, *46*(2), 245–254. https://doi.org/10.1016/j.nbd.2011.12.006

Koscianska, E., Kozlowska, E., & Fiszer, A. (2021). Regulatory Potential of Competing Endogenous RNAs in Myotonic Dystrophies. *International Journal of Molecular Sciences*, *22*(11), 6089. https://doi.org/10.3390/ijms22116089

Li, J., Wang, W., Xia, P., Wan, L., Zhang, L., Yu, L., … Xu, C. (2018). Identification of a five-lncRNA signature for predicting the risk of tumor recurrence in patients with breast cancer. *Identification of a Five-LncRNA Signature for Predicting the Risk of Tumor Recurrence in Patients with Breast Cancer*, *143*(9), 2150–2160. https://doi.org/10.1002/ijc.31573

Li, Z., Dou, P., Liu, T., & He, S. (2017). Application of Long Noncoding RNAs in Osteosarcoma: Biomarkers and Therapeutic Targets. *Cellular Physiology and Biochemistry*, *42*(4), 1407–1419. https://doi.org/10.1159/000479205

Marra, F., Lunetti, P., Curcio, R., Lasorsa, F. M., Capobianco, L., Porcelli, V., … Scarcia, P. (2021). An Overview of Mitochondrial Protein Defects in Neuromuscular Diseases. *Biomolecules*, *11*(11), 1633. https://doi.org/10.3390/biom11111633

Martínez-Rodríguez, J. E., Lin, L., Iranzo, A., Genis, D., Martí, M. J., Santamaria, J., & Mignot, E. (2003). Decreased Hypocretin-1 (Orexin-A) Levels in the Cerebrospinal Fluid of Patients with Myotonic Dystrophy and Excessive Daytime Sleepiness. *Sleep*, *26*(3), 287–290. https://doi.org/10.1093/sleep/26.3.287

Mizukami, K., Sasaki, M., Baba, A., Suzuki, T., & Shiraishi, H. (1999). An autopsy case of myotonic dystrophy with mental disorders and various neuropathologic features. *Psychiatry and Clinical Neurosciences*, *53*(1), 51–55. https://doi.org/10.1046/j.1440-1819.1999.00470.x

Nakamori, M., Sobczak, K., Puwanant, A., Welle, S., Eichinger, K., Pandya, S., … Thornton, C. A. (2013). Splicing biomarkers of disease severity in myotonic dystrophy. *Annals of Neurology*, *74*(6), 862–872. https://doi.org/10.1002/ana.23992

Otero, B. A., Kiril Poukalov, Hildebrandt, R. P., Thornton, C. A., Kenji Jinnai, Fujimura, H., … Wang, E. T. (2021). Transcriptome alterations in myotonic dystrophy frontal cortex. *Cell Reports*, *34*(3), 108634–108634. https://doi.org/10.1016/j.celrep.2020.108634

Palazzo, A. F., & Koonin, E. V. (2020). Functional Long Non-coding RNAs Evolve from Junk Transcripts. *Cell*. https://doi.org/10.1016/j.cell.2020.09.047

Prisiliana Velázquez-Bernardino, García-Sierra, F., Óscar Hernández-Hernández, Bermúdez, M., Geneviève Gourdon, Mário Gomes-Pereira, & Cisneros, B. (2011). Myotonic dystrophy type 1-associated CTG repeats disturb the expression and subcellular distribution of microtubule-associated proteins MAP1A, MAP2, and MAP6/STOP in PC12 cells. *Molecular Biology Reports*, *39*(1), 415–424. https://doi.org/10.1007/s11033-011-0753-y

Serra, L., Silvestri, G., Petrucci, A., Basile, B., Masciullo, M., Makovac, E., … Bozzali, M. (2014). Abnormal Functional Brain Connectivity and Personality Traits in Myotonic Dystrophy Type 1. *JAMA Neurology*, *71*(5), 603–611. https://doi.org/10.1001/jamaneurol.2014.130

Tahraoui-Bories, J., Mérien, A., Anchel González-Barriga, Lainé, J., Céline Leteur, Hélène Polvèche, … Martinat, C. (2023). MBNL-dependent impaired development within the neuromuscular system in myotonic dystrophy type 1. *Neuropathology and Applied Neurobiology*, *49*(1). https://doi.org/10.1111/nan.12876

Zhang, J., Zou, shuai, & Deng, L. (2018). Gene Ontology-based function prediction of long non-coding RNAs using bi-random walk. *BMC Medical Genomics*, *11*(S5). https://doi.org/10.1186/s12920-018-0414-2

2023 - 2024 Junior Science & Humanities Symposium
Statement of Outside Assistance

Students submitting their research paper to the Regional and National symposium must complete this form in full and submit with the final research paper.
Please type "N/A" in any field that is "not applicable" to your research.

Projects conducted without proper supervision will be disqualified from both regional and National competition. Further guidelines may be found at http://www.jshs.org & in the Core of Rules of Competition.

Student Participant to Complete:

Name:

Title of Paper:

School:

Teacher or Mentor Name:

Regional Symposium:

1. **Please explain your role in the development of the project idea.**

2. **What steps led you to formulate your research question?**
   – or – What steps led you to develop the design for your project?

3. **Where did you conduct the major part of your work? If an institutional setting, list the name.**
   (e.g., home, school, or other institutional setting such as university lab, medical center, etc.)

4. **Describe the assistance that you received throughout the project.**

5. **If you conducted your research in an institutional setting (i.e., university lab, medical center, etc.), describe your role on the team <u>AND</u> what role each person played in the research investigation?**

6. **Describe what parts of the research you did on your own and what parts where you received help.** *(e.g., literature search, hypothesis, experimental design, use of special equipment, gathering data, evaluation of data, statistical analysis, conclusions, and preparation of written report (abstract and/or paper))*

7. **Was any data set from an outside source (literature, handbooks, publicly available or privately shared tables, etc.) used in the research? If so, what data set(s) was used and from where was it sourced? What was its purpose in this research?**

8. **Was the data set sourced outside this research the only way to obtain the data required for the research? Please explain.**

9. **Is this research a continuation of an investigation that was previously submitted to a regional JSHS? If so, describe how you have expanded your investigation**.

10. **If this is a continuation project, please submit your previously submitted abstract here (copy/paste).**

11. **Did you utilize any form of AI in your research or paper? If so, explain what tools were used AND for what purpose.**

12. **Is there an IRB determination on file for this research? (Yes/No)**

13. **Provide the IRB determination details including IRB number, name, institution, and dates.**

14. **Provide full details about the processes and procedures utilized for your research, including usage & disposal of materials.**

Teacher and/or Supervising Mentor to Complete (or Parent if no teacher/mentor involvement):

Comments by teacher and/or supervising mentor on the students' individual contributions to the research investigation or engineering/computer science project. *If no Teacher or Mentor/Scientist was involved the **Parent** must complete this section describing their role in the research.*

When Tehreem joined her research team she was given a training module in which she was required to self teach basic statistics and computer coding, so that she could assist in her labs RNA work. Not only did Tehreem excel in this task, she also began taking online classes so that she could satisfy her own curiosity and prove to herself that she was ready to tackle her mentors investigations. It has been an incredible pleasure to watch Tehreem master and utilize all of her new skills for her project and all of it was done out of intense curiosity.

Statement by the teacher and/or supervising mentor acknowledging that the student conducted the research in accordance with proper procedures and protocols for the conduct of animal research or human research.

I, Nathaniel Covert, do atest that Tehreem Fatima conducted the research in accordance with proper procedures and protocols at the RNA Institute under the guidance of Dr. Andrew Berglund.

The Student, Teacher and/or Supervising Mentor must sign below. ***If you did the work without a teacher or supervising scientist, you will need a signature from your parent and a brief description of their role in the research.***

11/27/23	Tehreem Fatima	Shaker High School
Date	Signature of **Student** (required)	**Student's** High School
11/27/23		Nathaniel Covert
Date	Signature of **Teacher**	**Teacher's** High School
11/27/23		Dr. Andrew Berglund
Date	Signature of **Supervising Mentor**	Name of **Supervising Mentor**
	Director, Professor	RNA Insititute, University of Albany
	Title of **Supervising Mentor**	Institution of **Supervising Mentor**
11/27/23	Shazia Daud	Shazia Daud 7329258053
Date	Signature of **Parent** (required if no Teacher or Supervising Mentor was involved in the research.)	Name and Phone Number of **Parent**