Alternative Splicing in the Cardiac Tissue of Myotonic Dystrophy Type 1 Patients

Jared Richardson
Biology, Biochemistry

Faculty Mentor: Andy Berglund
Chemistry, Institute of Molecular Biology
Interim Assoc. Dean, Graduate School

Mentor: Adam Struck
Chemistry, Institute of Molecular Biology

Introduction

Myotonic dystrophy type 1 (DM1) is the most common form of adult-onset muscular dystrophy, affecting roughly 1/8000 people worldwide.[1] The disorder is an autosomal dominant RNA repeat disease caused by an expanded CTG repeat in the 3′ untranslated region (UTR) of the DMPK gene[1] (Figure 1). Transcription of the mutant DMPK forms a toxic RNA that is sequestered in the nucleus.[1] The toxic RNA, in turn, forms foci that accumulate in the nucleus and sequester different splicing regulators away from their normal targets, causing widespread splicing misregulation.[1] The MBNL proteins 1, 2, and 3 have been identified as splicing regulators especially prone to sequestration in DM1.[1] In addition to these sequestered proteins, another splicing regulator, CUGBP1, part of the CELF family, is hyperphosphoralated as a result of the expression of the toxic RNA repeats. The hyperphosphoration results in the stabilization of CUGBP1 and its presence at relatively higher levels in the tissues of DM1 patients. Together, the changes in the MBNL proteins and changes in the activities of the CELF protein family are thought to be the primary cause of the splic-

ing misregulation observed in DM1.[2] The missplicing creates many alternate protein products that may be either nonfunctional or vary from their normal function.[2] These proteins include INSR, which has been shown to cause insulin sensitivity in DM1 patients, and CLCN1, which has been shown to cause myotonia.[3] Other proteins including TNNT2 have been shown to cause cardiac defects in the disease.[4]

Previous researchers have not studied DM-related splicing changes on a global level in the hearts of human patients.[5] Instead, only a few missplicing events in cardiac muscle have been previously investigated, and not with the breadth that RNA-Seq on a transcriptome level could provide.[5] Studying the transcriptome, therefore, could provide much useful information on the missplicing events that occur in the cardiac muscle of DM1 patients. This is important because cardiac abnormalities account for the second highest cause of death in DM1 patients.[6]

In this study we compared the splicing patterns in cardiac muscle between DM1 patients and healthy controls. Numerous alternative exons were found to be differentially included in DM1 patients as compared to healthy controls. To investigate the potential affects these splicing changes may have on heart function, we examined the ontology of the parent genes. Previous research has suggested that a small list of these parent genes may have a causative effect in dilated cardiomyopathy (DCM).[7] These genes include TNNT2, TPM1, CAM2KD, TTN, VCL, MEF2D, LDB3, NEXN, OBSCN, SYN1, and PDLIM3.[8,9] These findings are important because prior research has already observed DCM in DM1 patients, and DCM is associated with a high mortality rate.[10,11] The missplicing of these genes may perhaps lead to the DCM observed in DM1 patients. This research may verify whether these genes are related to DCM and suggest additional avenues for future research.

Methods

The biological samples used in this study were previously taken by Tom Cooper, a lab collaborator at the Baylor College of Medicine, and then sequenced by Eric Want at the Massachusetts Institute of Technology using RNA-Seq. Total RNA was depleted of ribosomal RNA following the manufacturer's protocol and then prepared according to the manufacturer's protocol.[12] The resulting paired-end RNA-sequencing libraries were sequenced on an Illumina HiSeq 2000 to yield roughly 100 million 2 X 60bp reads per library. The number of samples totaled 12, all taken from the cardiac muscle of recently deceased people. Nine of these people were previously identified with DM1, and three were identified as not having DM1. The data that was used came in the form of raw sequencing data with a list noting which tags belonged to which sample. Library quality was verified using FastQC.[13] Next, the sequencing reads were compared to the human reference genome (version hg19) using GSNAP.[14] Isoform abundances were estimated, and each of the nine DM1 samples was compared to each of the three control samples using MISO (version 0.4.9).[15] A custom set of *hg19* alternative event annotations were generated from the Refseq, Ensembl, and UCSC gene models using custom scripts (available by request). For this analysis we focused only on the differential inclusion and exclusion of "cassette" exons between the DM1 and control samples.

The data were processed using multiple scripts in the graphic interface R Studio®, and the splicing events were organized based on whether the samples were from DM1 patients or from the controls. Following that step, one of the outputs from MISO, the percent spliced in (PSI), was averaged with a standard deviations of the grouped DM1 patients and the grouped controls. The averages of the PSI for the DM1 patients were subtracted from the controls for each splicing event. That result is referred to as the delta PSI and was the secondary factor for ranking the significance of each

event. The primary factor for ranking the significance of each event was another MISO output, the Bayes factor, which is a MISO measure of the confidence that the splicing events were actually observed in the samples. For example, a Bayes factor greater than five was considered to be significant. The comparison data was then filtered based on the criteria described below to identify misregulated splicing events.

Filtering criteria:

- At least ⅔ of the controls and DM1 samples had their PSI values retained for that event.
- The absolute value of the mean delta PSI for each DM1 sample and each control is greater than 0.05 (i.e., $|\overline{\Delta\psi}| \geq 0.05$).
- At least ¼ of DM1 patients showed sufficient evidence of differential splicing compared to the controls (median Bayes factor ≥ 5).

The parent genes of the filter splicing events were then matched to the ontology terms with which they were associated and subdivided into new lists based on terms related to their function. These functional terms included contraction, heart, muscle, heart contraction, and muscle contraction. Next, we searched Pubmed for information about genes included in the "heart contraction" and "muscle contraction" lists to see if those genes were related to any known diseases of the heart. Finally, a search of the significant event list used a list of genes previously found to be related to DCM to see if any of the genes were found on both lists.

Results

Using MISO, we quantified a total of more than 35,878 unique splicing events from all of our samples. The PSI values for each event were averaged for each group of samples and ranked by the number of comparisons between DM and control individuals if the Bayes factor was five or greater. After applying our filtering criteria, we found that 583 cassette

exons showed evidence of differential inclusion in heart tissue derived from DM1 patients relative to healthy control tissues. Of the parent genes for these events, four were associated with the term "heart contraction," 26 with the term "muscle contraction," 31 with the term "heart," 31 with the term "contraction," and 125 with the term "muscle." Three of the four genes associated with the term "heart contraction" were previously found to be associated with heart conditions (Figure 1). These genes included TNNT2, TPM1, and CAMK2D. Of the heart-related genes for which we searched on Pubmed, five genes of the 26 events related to the term "muscle contraction" were shown in previous research to be related to some kind of health condition of the heart. These genes included TNNT2, TPM1,[16] CAMK2D,[9] TTN,[17] and VCL.[7] By using the term "heart," we found one additional gene— MEF2D[18,19]—that previous research has seen as related to health conditions of the heart. We also found other genes in our list of 583 events that have been shown to be implicated in causing DCM.[7] These genes included LDB3, NEXN, OBSCN, and SYN1. With further research we also found one additional gene, PDLIM3,[20] to be related to DCM. This gave a total of 10 genes from 26 events from the list of 583 significant alternative splicing events that have been shown to be causative factors for DCM. The roles that each gene contributes to DCM are shown in Figure 1. A diagram of the roles of TNNT2, TTN, TPM1, LDB3, NEXN, VCL and OBSCN in the sarcomere is shown in Figure 2.

Discussion

The results showed that the ratio of mRNA splice isoforms is different in DM1 patients than in unaffected, healthy individuals. Altogether we identified 583 cassette exon events that passed our stringent cutoff guidelines and warranted study. We need to perform qRT-PCR to confirm that the splicing events are indeed observed at the rate seen in our study. Of

GENE	Protein	Protein Contribution to DCM	Heart Related Illness
TNNT2	Troponin T type 2[27]	Sarcomere[27]	Dilated Cardiomyopathy, other cardiomyopathies[27]
TPM1	Tropomyosin 1[27]	Sarcomere[27]	Dilated Cardiomyopathy, other cardiomyopathies[27]
TTN	Titin[27]	Sarcomere[27]	Dilated Cardiomyopathy, other cardiomyopathies[27]
VCL	Vinculin[27]	Z-disk[27]	Dilated Cardiomyopathy, other cardiomyopathies[27]
MEF2D	Myocyte Enhancer Factor 2D[18]	Transcription enhancer[18]	Stress dependent remodeling of heart[18]
LDB3	LIM domain binding 3[27]	Z-disk[27]	Dilated Cardiomyopathy, other cardiomyopathies[27]
NEXN	Nexilin (F actin-binding protein)[27]	Z-disk[27]	Dilated Cardiomyopathy, other cardiomyopathies[27]
OBSCN	Obscurin, Cytoskeletal Calmodulin And Titin-Interacting RhoGE[7]	Protein Kinase[7]	Dilated Cardiomyopathy[7]
SYN1	Synapsin 1[7]	Nuclear membrane[7]	Dilated Cardiomyopathy[7]
PDLIM3	PDZ and LIM domain 3[27]	Cytoskeleton[27]	Dilated Cardiomyopathy[27]
CAM2KD	Calcium/Calmodulin-Dependent Protein Kinase Type II Delta Chain[9]	Protein Kinase[9]	Dilated Cardiomyopathy[9]

Figure 1: Table of Genes Found in DCM and DM1.

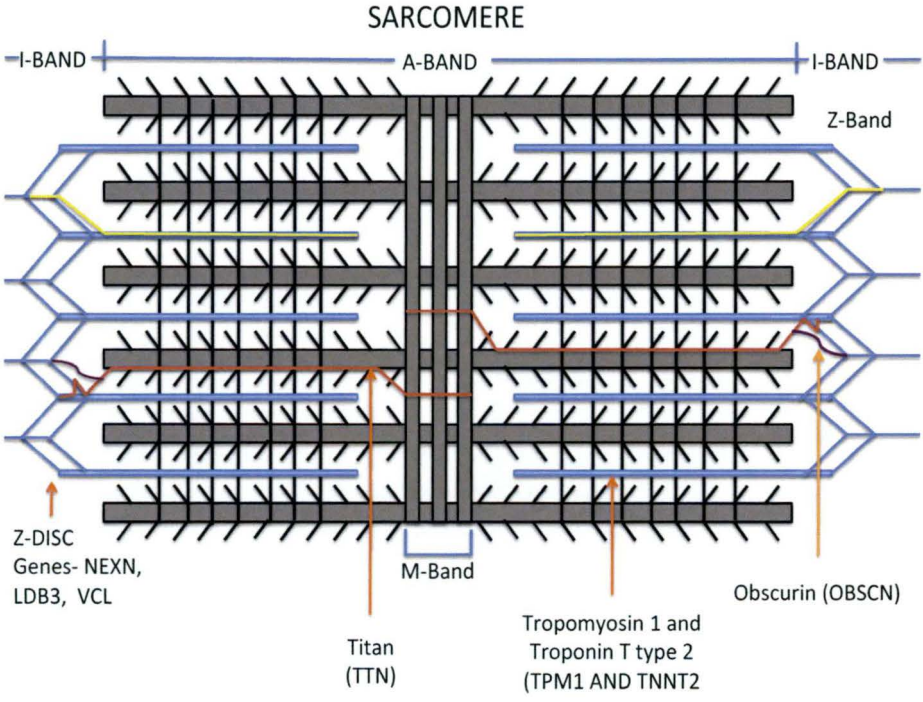

Figure 2: Genes and Role in Sarcomere. The genes NEXN, LDB3, and VCL make up proteins that are part of the Z-disc. Titan is a large protein that helps anchor the M-Band, A-Band, and I-Band together. Obscurin helps anchor Titan to the Z-band. Tropomyosin 1 is a coiled protein and helps regulate contraction. Troponin T type 2 helps regulate calcium.

these 583 alternate splicing events, we found that many were from genes related to vital functions of the heart with the most important and descriptive terms being "muscle contraction," "heart," and "heart contraction." Of all the genes on both lists, we found six that previous researchers found to be related to significant health conditions of the heart. We also found an additional seven genes of the 583 significant splicing events that have been shown to have mutations that can cause DCM. This gives a total of

10 genes that have been shown to have a causative effect in DCM and, based on this study, show a significant likelihood of being misspliced. Because MBNL is important in development and regulates the transition from fetal to adult isoforms, the genes seen in this study could be reverting to their fetal isoforms. If this is occurring, the protein products of these isoforms may not be able to function properly, thereby leading to DCM. This potential is important because DCM commonly causes heart failure and sudden cardiac death.[21] This finding is of even greater importance in DM1 because the second leading cause of death in DM1 patients relates to cardiac abnormalities.

Dilated cardiomyopathy has been observed in DM1 patients before.[10] In one case study that examined the autopsies of 12 DM1 patients, 25% of patients had obvious signs of DCM.[24] Although the degree of prevalence of DCM in DM1 patients is not yet known, one of the largest cross sectional studies of heart-related illnesses in DM1 patients found that structural abnormalities were present in the form of left ventricular systolic dysfunction (20.6%) and reduced global longitudinal strain (21.7%).[25] Both abnormalities are signs of DCM,[26,27] although DCM was not specifically examined in that cross sectional study. The missplicing of MEF2D, a transcription factor that has been shown to mediate stress dependent remodeling in the heart,[18] could possibly add to the effect of DCM in DM1. If the missplicing of the MEF2D gene inhibits this protein's function, then it will be unable to mediate the stress caused by DCM, which could lead to more severe clinical effects of DCM.

Most of the genes found in relation to DCM and the genes found in this study have some function in the sarcomere of muscle[7] (Figure 2). In past studies of the genes in DCM, researchers observed mutations that inhibited the proper functions related to muscle contraction.[7] Given the large number of mutations that contribute to DCM, the pathophysiology is

complex and not completely understood.[26] Many mutations have been observed in genes that directly contribute to DCM through truncations of the protein product.[27] The most investigated genes include TTN, TNNT2, and TPM1.[7] In both TTN and TNNT2, some of the mutations that contribute to DCM include mutations that cause missplicing.[28] Therefore, it would be possible for missplicing in at least two of these genes to lead to DCM. Whether the same missplicing events seen in DM1 can contribute to the same missplicing seen in these genes in some cases of DCM requires future study.

The results of this research suggest many possible areas of future research. For example, researchers could investigate the functional role of the 10 genes misspliced in DM1 that are also implicated in other heart conditions. To do so, it would be important to confirm these splicing events using qRT-PCR. It could also prove useful for future studies to see if any other kinds of splicing events could contribute to DCM as our study examined only one kind of missplicing event. It will also be important to study the functions of the aberrant isoforms expressed as a result of the missplicing in DM1 to understand the consequences of this misregulation. All of the genes under study were shown to have at least one of these events, and many had far more. If the function of these genes is inhibited in the same way as the mutations seen in these genes in DCM, the misfunction of those genes could perhaps cause DCM. We will have to test these genes to see if the observed alternative splicing events could mimic these mutations and if the rates of alternative splicing are significant enough to do so as well. By completing these tests, we may actually be able to see if any of these genes with the splicing seen in DM1 are responsible for DCM in DM1 patients and perhaps their premature death.

References

1. Lee, J. E.; Cooper, T. A. Pathogenic Mechanisms of Myotonic Dystrophy. *Biochem. Soc. Trans.* **37,** (2009).
2. Meola, G. Clinical Aspects, Molecular Pathomechanisms and Management of Myotonic Dystrophies. *Acta Myol.* **32,** 154–165 (2013).
3. Kino, Y. *et al.* MBNL and CELF Proteins Regulate Alternative Splicing of the Skeletal Muscle Chloride Channel CLCN1. *Nucleic Acids Res.* **37,** 6477–6490 (2009).
4. Kuyumcu-Martinez, N. M.; Cooper, T. A. Misregulation of Alternative Splicing Causes Pathogenesis in Myotonic Dystrophy. *Prog. Mol. Subcell. Biol.* **44,** 133–159 (2006).
5. GAO, C.; Wang, Y. Global Impact of RNA Splicing on Transcriptome Remodeling in the Heart. *J. Zhejiang Univ. Sci. B* **13,** 603–608 (2012).
6. Mankodi, A.; Lin, X.; Blaxall, B. C.; Swanson, M. S.; Thornton, C. A. Nuclear RNA Foci in the Heart in Myotonic Dystrophy. *Circ. Res.* **97,** 1152–1155 (2005).
7. McNally, E. M.; Golbus, J. R.; Puckelwartz, M. J. Genetic Mutations and Mechanisms in Dilated Cardiomyopathy. *J. Clin. Invest.* **123,** 19–26 (2013).
8. Hershberger, R. E.; Morales, A.; Siegfried, J. D. Clinical and Genetic Issues in Dilated Cardiomyopathy: A review for Genetics Professionals. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **12,** 655–667 (2010).
9. Toko, H. *et al.* Ca^{2+}/Calmodulin-dependent Kinase IIΔ Causes Heart Failure by Accumulation of p53 in Dilated Cardiomyopathy. *Circulation* **122,** 891–899 (2010).
10. Schoser, B.; Timchenko, L. Myotonic Dystrophies 1 and 2: Complex Diseases with Complex Mechanisms. *Curr. Genomics* **11,** 77–90 (2010).
11. Taylor, M. R.; Carniel, E.; Mestroni, L. Cardiomyopathy, Familial Dilated. *Orphanet J. Rare Dis.* **1,** 27 (2006).
12. Cheranova, D. *et al.* RNA-seq Analysis of Transcriptomes in Thrombin-treated and Control Human Pulmonary Microvascular Endothelial Cells. *J. Vis. Exp. JoVE* (2013). doi:10.3791/4393
13. Andrews, S. FastQC: A Quality Control tool for High Throughput Sequence Data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

14. Wu, T. D.; Nacu, S. Fast and SNP-tolerant Detection of Complex Variants and Splicing in Short Reads. *Bioinformatics* **26,** 873–881 (2010).

15. Katz, Y.; Wang, E. T.; Airoldi, E. M.; Burge, C. B. Analysis and Design of RNA Sequencing Experiments for Identifying Isoform Regulation. *Nat. Methods* **7,** 1009–1015 (2010).

16. Nonaka, M.; Morimoto, S. Experimental Models of Inherited Cardiomyopathy and its Therapeutics. *World J. Cardiol.* **6,** 1245–1251 (2014).

17. Mestroni, L.; Taylor, M. R. G. Genetics and Genetic Testing of Dilated Cardiomyopathy: A New Perspective. *Discov. Med.* **15,** 43–49 (2013).

18. Kim, Y. *et al.* The MEF2D Transcription Factor Mediates Stress-dependent Cardiac Remodeling in Mice. *J. Clin. Invest.* **118,** 124–132 (2008).

19. Fielitz, J. *et al.* Requirement of Protein Kinase D1 for Pathological Cardiac Remodeling. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 3059–3063 (2008).

20. Hershberger, R. E.; Hedges, D. J.; Morales, A. Dilated Cardiomyopathy: The Complexity of a Diverse Genetic Architecture. *Nat. Rev. Cardiol.* **10,** 531+ (2013).

21. Lakdawala, N. K.; Winterfield, J. R.; Funke, B. H. Dilated Cardiomyopathy. *Circ. Arrhythm. Electrophysiol.* **6,** 228–237 (2013).

22. Koshelev, M.; Sarma, S.; Price, R. E.; Wehrens, X. H. T.; Cooper, T. A. Heart-specific Overexpression of CUGBP1 Reproduces Functional and Molecular Abnormalities of Myotonic Dystrophy Type 1. *Hum. Mol. Genet.* **19,** 1066–1075 (2010).

23. Wang, G.-S.; Kearney, D. L.; De Biasi, M.; Taffet, G.; Cooper, T. A. Elevation of RNA-binding Protein CUGBP1 is an Early Event in an Inducible Heart-specific Mouse Model of Myotonic Dystrophy. *J. Clin. Invest.* **117,** 2802–2811 (2007).

24. Lin, A. E.; Mitchell, F. M.; Fitz, R. W.; Doyle, J. J. Dilated Cardiomyopathy in Myotonic Dystrophy. *J. Am. Coll. Cardiol.* **13,** 262–263 (1989).

25. Petri, H. *et al.* High Prevalence of Cardiac Involvement in Patients with Myotonic Dystrophy Type 1: A Cross-sectional Study. *Int. J. Cardiol.* **174,** 31–36 (2014).

26. Duan, F. *et al.* Preliminary Clinical Study of Left Ventricular Myocardial Strain in Patients with Non-ischemic Dilated Cardiomyopathy by Three-dimensional Speckle Tracking Imaging. *Cardiovasc. Ultrasound* **10,** 8 (2012).

27. Gopal, D. M.; Sam, F. New and Emerging Biomarkers in Left Ventricular Systolic Dysfunction—Insight into Dilated Cardiomyopathy. *J Cardiovasc. Transl. Res.* **6,** 516–527 (2013).

28. Kong, S. W. *et al.* Heart Failure Associated Changes in RNA Splicing of Sarcomere Genes. *Circ. Cardiovasc. Genet.* **3,** 138–146 (2010).