

## GENERAL ARTICLE

# Transcriptome alterations in myotonic dystrophy skeletal muscle and heart

Eric T. Wang<sup>1,2,3,4,5,\*</sup>, Daniel Treacy<sup>1</sup>, Katy Eichinger<sup>6</sup>, Adam Struck<sup>11</sup>, Joseph Estabrook<sup>1,2,3,4,5</sup>, Hailey Olafson<sup>3,4,5</sup>, Thomas T. Wang<sup>1</sup>, Kirti Bhatt<sup>6</sup>, Tony Westbrook<sup>12</sup>, Sam Sedehizadeh<sup>12</sup>, Amanda Ward<sup>7</sup>, John Day<sup>10</sup>, David Brook<sup>12</sup>, J. Andrew Berglund<sup>3,4,5,11</sup>, Thomas Cooper<sup>7,8,9</sup>, David Housman<sup>1,2</sup>, Charles Thornton<sup>6</sup> and Christopher Burge<sup>1,2</sup>

<sup>1</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA 02139, <sup>2</sup>Koch Institute, Massachusetts Institute of Technology, Cambridge, MA, USA 02139, <sup>3</sup>Department of Molecular Genetics & Microbiology, University of Florida, Gainesville, FL, USA 32610, <sup>4</sup>Center for NeuroGenetics, University of Florida, Gainesville, FL, USA 32610, <sup>5</sup>University of Florida Genetics Institute, University of Florida, Gainesville, FL, USA 32610, <sup>6</sup>Department of Neurology, University of Rochester, Rochester, NY, USA 14627, <sup>7</sup>Department of Pathology & Immunology, Baylor College of Medicine, Houston, TX, USA 77030, <sup>8</sup>Department of Molecular Physiology & Biophysics, Baylor College of Medicine, Houston, TX, USA 77030, <sup>9</sup>Department of Molecular & Cellular Biology, Baylor College of Medicine, Houston, TX, USA 77030, <sup>10</sup>Department of Neurology, Stanford University, Palo Alto, CA, USA 94304, <sup>11</sup>Department of Biochemistry, University of Oregon, Eugene, OR, USA 97403 and <sup>12</sup>School of Life Sciences, Queen's Medical Center, University of Nottingham, Nottingham, UK

\*To whom correspondence should be addressed at: Department of Molecular Genetics & Microbiology, University of Florida Gainesville, FL USA 32610. Tel: +352-273-7601; E-mail: eric.t.wang@ufl.edu

## Abstract

Myotonic dystrophy (*dystrophia myotonica*, DM) is a multi-systemic disease caused by expanded CTG or CCTG microsatellite repeats. Characterized by symptoms in muscle, heart and central nervous system, among others, it is one of the most variable diseases known. A major pathogenic event in DM is the sequestration of muscleblind-like proteins by CUG or CCUG repeat-containing RNAs transcribed from expanded repeats, and differences in the extent of MBNL sequestration dependent on repeat length and expression level may account for some portion of the variability. However, many other cellular pathways are reported to be perturbed in DM, and the severity of specific disease symptoms varies among individuals. To help understand this variability and facilitate research into DM, we generated 120 RNASeq transcriptomes from skeletal and heart muscle derived from healthy and DM1 biopsies and autopsies. A limited number of DM2 and Duchenne muscular dystrophy samples were also sequenced. We analyzed splicing and gene expression, identified tissue-specific changes in RNA processing and uncovered transcriptome changes strongly correlating with muscle strength. We created a web resource at <http://DMseq.org> that hosts raw and processed transcriptome data and provides a lightweight, responsive interface that enables browsing of processed data across the genome.

Received: July 30, 2018. Revised: November 30, 2018. Accepted: December 10, 2018

© The Author(s) 2018. Published by Oxford University Press. All rights reserved.

For Permissions, please email: journals.permissions@oup.com

## Introduction

Myotonic dystrophy types 1 and 2 (*dystrophia myotonica*, DM1 and DM2) are microsatellite repeat diseases caused by expansions of CTG or CCTG repeats in the *dystrophia myotonica* protein kinase (DMPK) (1) or cellular nucleic acid-binding protein (CNBP) (2) genes, respectively. Both DM1 and DM2 are highly multi-systemic, with disease symptoms occurring in skeletal muscle, cardiac muscle, the central nervous system, gastrointestinal tract, endocrine system and others. Symptoms including myotonia, muscle weakness, cardiac arrhythmia and hypersomnolence affect many patients, but the timing and severity of each symptom can vary dramatically between individuals and even across members of the same family.

Multiple molecular mechanisms are proposed to cause DM symptoms, including titration of muscleblind-like (MBNL) RNA-binding proteins by toxic, expanded CUG/CCUG repeat-containing transcripts that originate from the DMPK or CNBP loci [reviewed in (3)]; elevated levels of CUGBP/ETR-like factors (CELFs) due to hyperactivation of protein kinase C (4); microRNA dysregulation (5–7); alterations to signaling pathways such as GSK3 $\beta$  (8), PDGFR $\beta$ , PI3K/AKT and AMPK (9,10); repeat-associated non-ATG-dependent (RAN) translation of transcripts generated from repeat loci (11); and aberrant methylation of the expanded repeat tract (12). While all of these pathways may contribute to DM phenotypes, the best characterized pathway is sequestration of muscleblind-like (MBNL) proteins by RNAs containing expanded CUG or CCUG repeats, transcribed from DMPK or CNBP loci, respectively. Functional depletion of MBNLs causes changes in alternative splicing (13), alternative polyadenylation (14), RNA localization (15) and microRNA processing (6), and multiple MBNL-mediated mis-splicing events have been linked to specific symptoms. These events include CLCN1 exon 7a (myotonia) (16,17), BIN1 exon 11 (muscle weakness) (18), exon 22 of CACNA1S (muscle weakness) (19) and exon 5A of SCN5A (cardiac arrhythmia) (20). Accordingly, the extent of mis-splicing of MBNL targets correlates with muscle strength and has formed the basis for multiple efforts to develop biomarkers of disease severity (21–23).

Although MBNL depletion is a major pathogenic event in DM, many additional pathways may play important roles in the disease. As MBNL itself is a major regulator of RNA processing, some variability in DM could stem from polymorphisms in cis-elements that regulate the repertoire of MBNL splicing targets, or from these additional pathways. Furthermore, most work to date has focused on skeletal muscle pathology, and our knowledge of molecular changes in other muscle types such as cardiac muscle is less extensive. Unbiased studies of transcriptomes derived from these tissues could facilitate studies to better understand all of these areas. Here, we describe a series of RNA-Seq transcriptomes that we have generated from autopsies and biopsies of skeletal and heart muscle of subjects affected by DM1, DM2 and unrelated muscular dystrophies, as well as healthy unaffected individuals. Transcriptome signatures were found to clearly separate DM from non-DM samples when analyzing either alternative splicing or gene expression changes. These transcriptomes are available as a public resource at GEO (GSE86356), as well as through a web server (<http://DMseq.org>) that provides an interface to view sequence read density across genomic loci. This web server also provides tools to easily view gene expression and isoform abundance estimates and will facilitate the development and exploration of new hypotheses about disease pathogenesis in DM, as well as biomarkers for use in clinical trials.

## Results

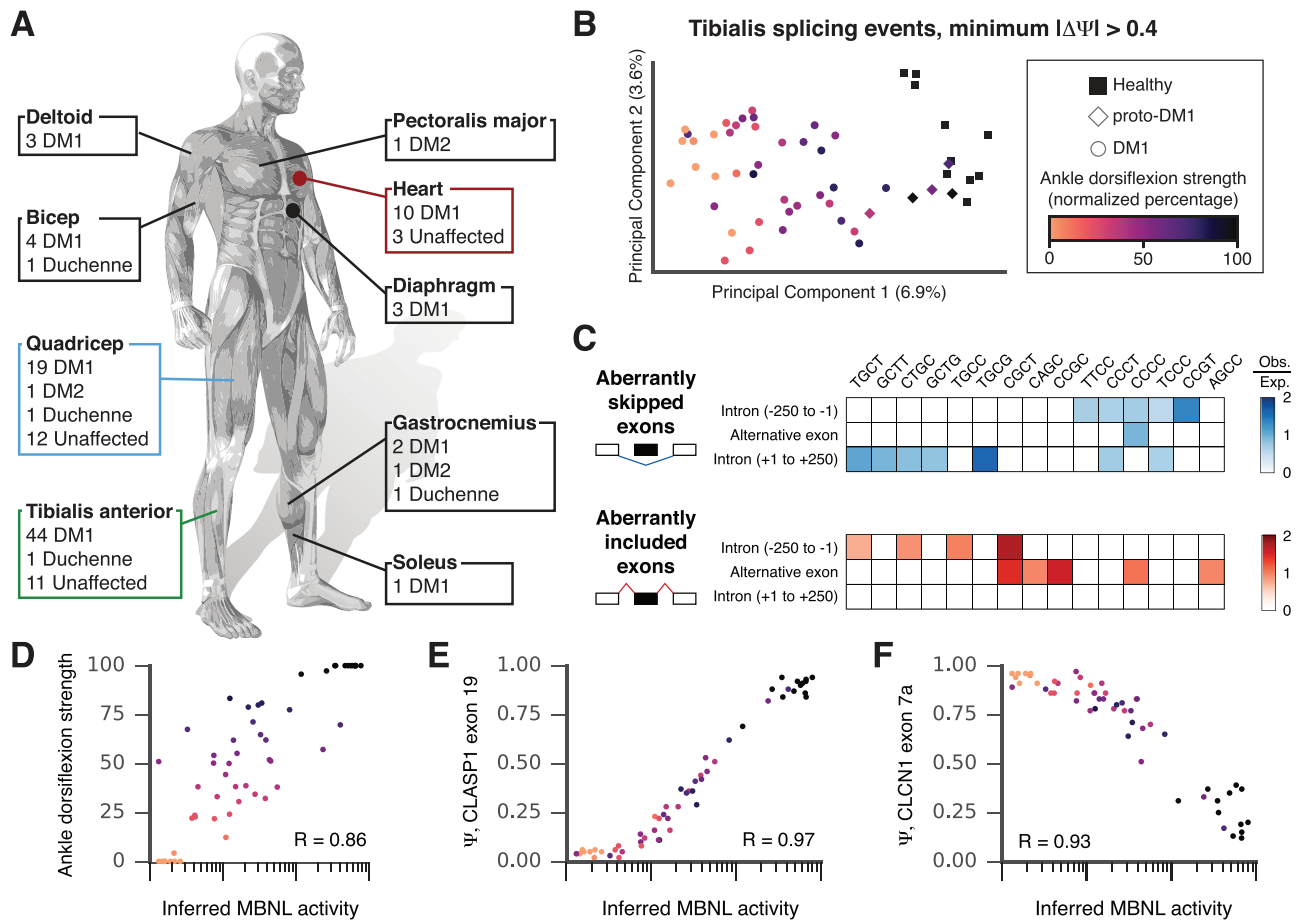
### Biopsy and autopsy muscle and heart tissue samples were subjected to RNA-Seq

We performed RNA-Seq on a series of heart and muscle autopsies and biopsies, from various regions of the body (Fig. 1A). These included a series of 55 tibialis anterior biopsies, including 44 DM1 subjects and 11 unaffected individuals, a series of 19 DM1 quadriceps biopsies and 12 unaffected individuals and a series of 10 DM1 heart autopsies and 3 unaffected individuals. Autopsy material was also obtained such that multiple muscle groups from the same individual could be compared with each other. For example, a single DM2 individual provided 5 distinct muscle types and 5 individuals with DM1 provided 2 or more muscle groups. All samples were sequenced to a depth of at least 41 M reads, with a subset of 50 samples subjected to extremely deep sequencing (>200 M reads), facilitating accurate assessment of gene expression and splicing for a major portion of the transcriptome (Supplementary Material, Table S1).

### A splicing signature for DM1

Because a role for MBNL-mediated mis-splicing is well established in DM1, we first investigated the extent to which a splicing signature could be observed across all DM1 tibialis biopsies. These DM1 tibialis biopsies were obtained from individuals who had also undergone quantitative muscle strength testing, including measurements of ankle dorsiflexion strength (Supplementary Material, Table S2). We computed percent spliced in (PSI or  $\Psi$ ) using MISO (24) for all samples (Supplementary Material, Table S3) and identified exons showing large differences in  $\Psi$ , with  $\Delta\Psi > 0.4$  (see Materials and Methods). Principal component analysis was performed on all samples using these exons and plotted for the first two principal components (Fig. 1B). As expected, healthy individuals clustered together, with proto-DM1 samples (fewer than 90 CTG repeats) also clustering nearby. DM1 tibialis samples were distributed across the first principal component in a manner that correlated with normalized ankle dorsiflexion strength (see Materials and Methods).

Previously, we had demonstrated that mis-splicing across many MBNL-dependent exons could be used to predict underlying levels of MBNL protein activity, which in turn might reflect a combination of total CUG repeat load and baseline MBNL levels unique to each individual (22). In previous studies, we observed that intracellular free (i.e. non-sequestered) MBNL protein concentration could be accurately inferred using a subset of MBNL-dependent splicing events in HEK293 cells. We applied this method to estimate free MBNL concentration—referred to here as ‘Inferred MBNL activity’—in DM1 tibialis biopsies. We again related all cassette exon  $\Psi$  values to inferred MBNL activity levels using sigmoid curve fits and focused subsequent analyses on the best-fitting events (see Materials and Methods). Four base motifs were enumerated within intronic sequences flanking these cassette exons (–250 to –1 and +1 to +250) and within the cassette exons themselves and compared to poorly fitting exons. The ratio of observed versus expected motif occurrences (Fig. 1C) revealed a strong signature for motifs known to interact with MBNL proteins (e.g. YGCY motifs, where Y = C or U) (25) upstream of and within aberrantly included exons and downstream of aberrantly skipped exons. Interestingly, a strong signature for pyrimidine-rich motifs was also identified upstream of aberrantly skipped exons. We found that inferred



**Figure 1.** All muscle transcriptomes generated and analysis of splicing changes in DM1 tibialis biopsies. (A) List of all tissues and samples subjected to RNAseq. Biopsy samples include all DM1 and unaffected tibialis samples, 12 DM1 quadriceps samples and 6 unaffected quadriceps samples. All other samples are autopsies. (B) Principal component analysis of tibialis biopsies using a subset of splicing events that show a minimum change of  $|\Delta\Psi| > 0.4$  between any pair of samples. Each point is a single tibialis biopsy from a single individual; squares show healthy individuals, diamonds show proto-DM1 individuals and circles show DM1 individuals. Points are also colored by normalized ankle dorsiflexion strength. (C) Motif analysis was performed on a set of high-confidence exons aberrantly skipped or included in DM1 tibialis biopsies (see [Materials and Methods](#)). The 4-mers were counted in the 250 nucleotides upstream, downstream and within alternative exons and compared to control exons that do not change in DM1. The heatmap shows colors for motifs that are significant ( $P < 1e-5$ , Bonferroni corrected). (D) The concentration of MBNL as inferred by splicing correlated to normalized ankle dorsiflexion strength. Points are also colored by strength according to the legend in (B). (E)  $\Psi$  for CLASP1 exon 19 as a function of (MBNL) inferred from splicing. (F)  $\Psi$  for CLCN1 exon 7a as a function of (MBNL), inferred from splicing.  $P < 1e-15$  for correlations in (D)–(F).

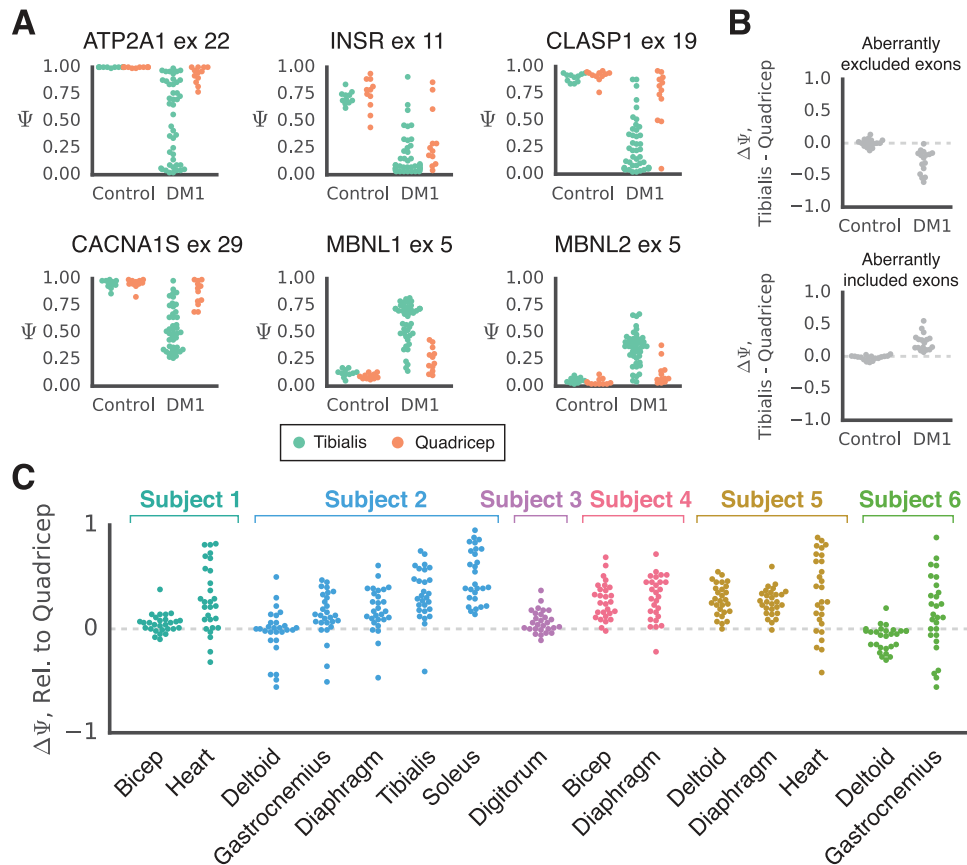
MBNL activity correlated strongly with  $\Psi$  for specific exons, such CLASP1 exon 19 and CLCN1 exon 7a, and also correlated strongly with normalized ankle dorsiflexion strength ( $R = 0.86$ ; [Fig. 1D–F](#)). These observations support the concepts that DM1 muscle from different subjects exhibits a continuum of severity, reflected by splicing dysregulation as well as muscle strength and that sensitive measurements of either of these parameters might allow for quantitative assessment of therapeutic interventions (21). Because biochemical estimates of the amount of MBNL protein depleted in any given tibialis biopsy are not available, it is not possible at this time to assign precise MBNL protein concentrations to specific levels of activity in this tissue ([Supplementary Material, Fig. S1A](#)).

Although numerous previous analyses of isoform alterations have focused on splicing events, MBNL proteins also regulate alternative polyadenylation (14). We therefore assessed changes in alternative last exon usage, an isoform change that can be controlled by splicing and/or alternative polyadenylation, often resulting in distinct protein C-termini. We found that inclusion of many alternative last exons strongly correlated

with estimated MBNL activity as well as muscle strength ([Supplementary Materials, Table S4 and Fig. S1B](#)). The functions of many of these exons remain to be elucidated. Similar to cassette exon splicing, the ability to measure distinct isoforms from the same gene makes them convenient as biomarkers that may be relatively robust despite potential differences in cell type composition of biopsy tissue.

### Splicing analyses across multiple DM1 muscles

Not all skeletal muscles are equally affected in DM1. For example, finger flexors, the tibialis anterior and neck muscles tend to be impacted earlier than other muscle groups and there is a distal to proximal pattern of muscle weakness and wasting. In DM2, a different pattern is observed as proximal muscles tend to be more severely affected. Here, we compared splicing dysregulation across distinct muscle groups, some obtained via biopsy and others via autopsy. First,  $\Psi$  values were computed for quadriceps biopsies obtained from 11 healthy subjects

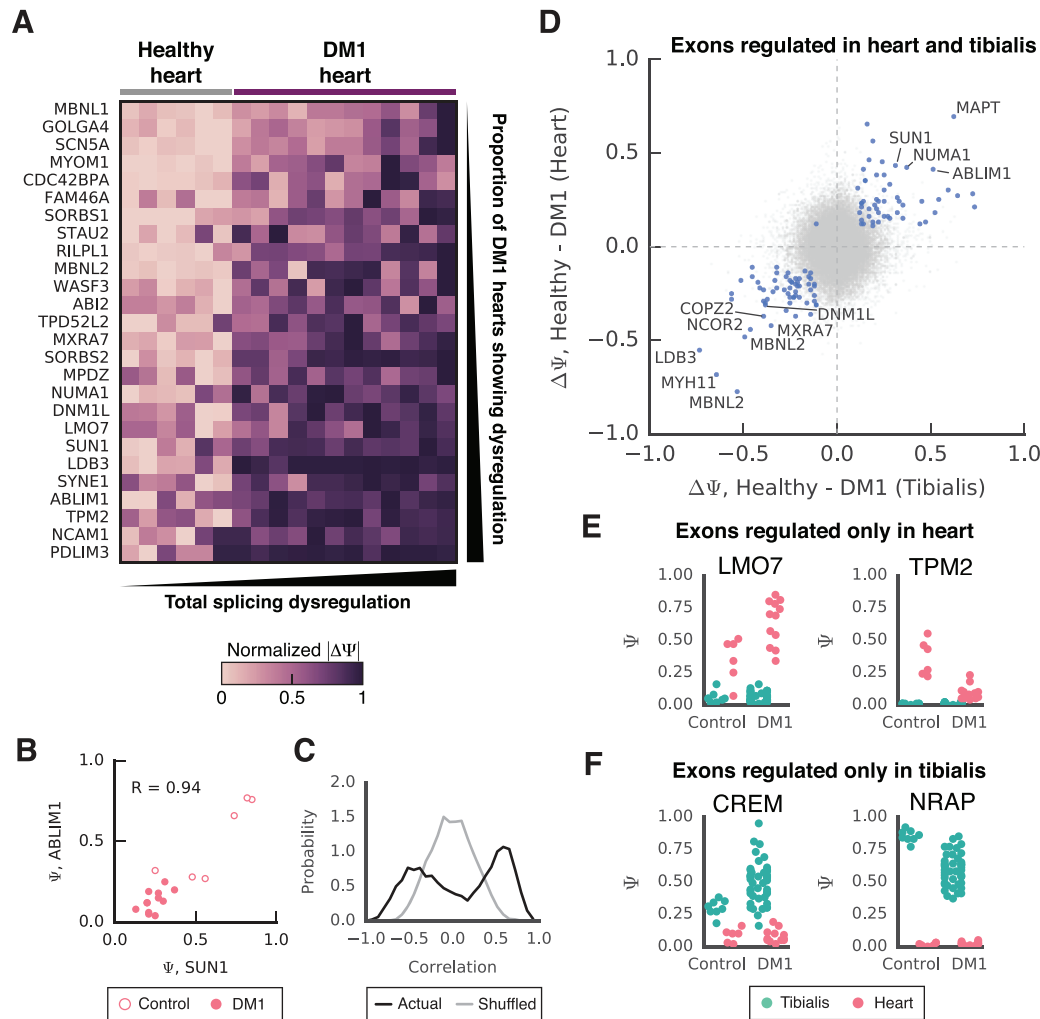


**Figure 2.** Analyses of splicing changes across distinct DM1 skeletal muscles. (A)  $\Psi$  for selected alternative exons from unaffected and DM1 tibialis and quadriceps biopsies. Tibialis  $\Psi$  values are shown in teal and quadriceps  $\Psi$  values in orange. (B) A set of high-confidence exons (see [Materials and Methods](#)) were separated into those aberrantly excluded (top) or included (bottom) in DM1. The median  $\Psi$  for each exon was then obtained for four groups—control quadriceps, control tibialis, DM1 quadriceps and DM1 tibialis. The difference in median  $\Psi$  ( $\Delta\Psi$ ) for each exon was computed between quadriceps and tibialis, separately for control and DM1 biopsies, and plotted. Each dot represents the difference in medians for a single exon. In general, exons are more severely dysregulated in tibialis than in quadriceps. (C) Similar to (B),  $\Psi$  values for a subset of exons (see [Materials and Methods](#)) were computed for several skeletal muscles obtained from the same autopsy subjects, all of whom contributed quadriceps. The difference in  $\Psi$  ( $\Delta\Psi$ ) relative to quadriceps was plotted for each exon. Here, aberrantly included exons and aberrantly excluded exons were plotted together, but the mis-splicing value for aberrantly excluded exons was multiplied by  $-1$  so that exons less severely affected in quadriceps are above the  $\Delta\Psi = 0$  line. In most cases, quadriceps exhibits the most mild splicing dysregulation relative to all other skeletal muscles analyzed.

and 12 DM1 subjects and compared to the tibialis  $\Psi$  values described above. Transcriptome-wide analyses showed that  $\Psi$  in healthy quadriceps correlated strongly with  $\Psi$  in healthy tibialis ([Supplementary Material, Fig. S2A](#)). However, as typified by six well-established MBNL targets, and as observed previously (21), the extent of mis-splicing was much more severe in DM1 tibialis relative to DM1 quadriceps ([Fig. 2A](#)). The median  $\Psi$  within each cohort (healthy quadriceps, healthy tibialis, DM1 quadriceps, DM1 tibialis) was computed for each exon, and the median quadriceps  $\Psi$  value was subtracted from the median tibialis  $\Psi$  value for the healthy cohorts and DM1 cohorts ([Fig. 2B](#)). On a per-exon basis, exons aberrantly excluded in DM1 were more strongly excluded in tibialis relative to quadriceps, and exons aberrantly included in DM1 were more strongly included in tibialis relative to quadriceps. These observations suggest that tibialis has more severe splicing dysregulation than quadriceps in DM. However, a caveat of this analysis is that the tibialis and quadriceps biopsies were obtained from separate cohorts, which might differ in disease severity. To further explore this hypothesis, we turned to our autopsy transcriptomes, some obtained from multiple skeletal muscles of the same individual. We

computed  $\Psi$  for quadriceps and at least one other muscle from six individuals, and  $\Delta\Psi$  relative to quadriceps was plotted for each additional muscle type ([Fig. 2C](#)). For ease of interpretation,  $\Delta\Psi$  values for exons aberrantly excluded in DM1 were multiplied by  $-1$  such that exons lying above  $\Delta\Psi = 0$  are those exhibiting less severe dysregulation in quadriceps. With the exception of deltoids from subjects 2 and 6, the muscle type exhibiting least severe splicing dysregulation was quadriceps. Interestingly, the level of DMPK mRNA was not significantly different in healthy quadriceps as compared to healthy tibialis, but the levels of MBNL1 mRNA are  $\sim 20\%$  higher in healthy quadriceps as compared to healthy tibialis ([Supplementary Material, Fig. S2B](#)).

We also analyzed splicing patterns across a limited number of DM2 and Duchenne muscular dystrophy (DMD) autopsy samples.  $\Psi$  was quantitated in four muscle groups from a single DM2 individual and four muscle groups from a single DMD individual. While quadriceps tended to be the most mildly affected muscle among those tested in DM1, this was not the case in DM2 muscles ([Supplementary Material, Fig. S2C](#)). Some splicing events were dysregulated in DMD muscle in a manner similar to that observed in DM1 (e.g. PKM, MEF2C, MEF2D;



**Figure 3.** Analyses of splicing changes in the DM1 heart. (A) A heatmap showing normalized splicing dysregulation ( $\Delta\Psi$ ) for the most severely dysregulated exons are displayed, computed from heart autopsy transcriptomes. The darker the color, the stronger the dysregulation. Individual hearts were sorted by the total dysregulation of splicing across multiple events, and splicing events were sorted by the proportion of DM1 hearts showing dysregulation (see [Materials and Methods](#)). (B) Scatter plot of  $\Psi$  for ABLIM1 exon 11 and SUN1 exon 3 across heart autopsy samples. Open circles are control hearts and filled-in circles are DM1 hearts. Pearson correlation value is also shown. (C) Histogram of correlation coefficients for all pairwise comparisons of  $\Psi$  for a set of high-confidence exons dysregulated in DM1 heart are shown in black (see [Materials and Methods](#)). A similar histogram computed using shuffled sample identities is illustrated for comparison (gray). (D) Scatter plot of  $\Delta\Psi$  in heart versus tibialis, where  $\Delta\Psi$  was obtained by subtracting the median DM1  $\Psi$  from the median control  $\Psi$  value (11 control tibialis biopsies, 44 DM1 tibialis biopsies, 6 control hearts, 12 DM1 hearts). The most strongly dysregulated exons are labeled. (E)  $\Psi$  values for LMO7 exon 20 and TPM2 mutually exclusive exon 6A (rel. to 6B), both dysregulated in DM1 heart but not tibialis. (F)  $\Psi$  values for CREM mutually exclusive exon 1a (rel. to 1b) and NRAP exon 12, both dysregulated in DM1 tibialis but not heart.

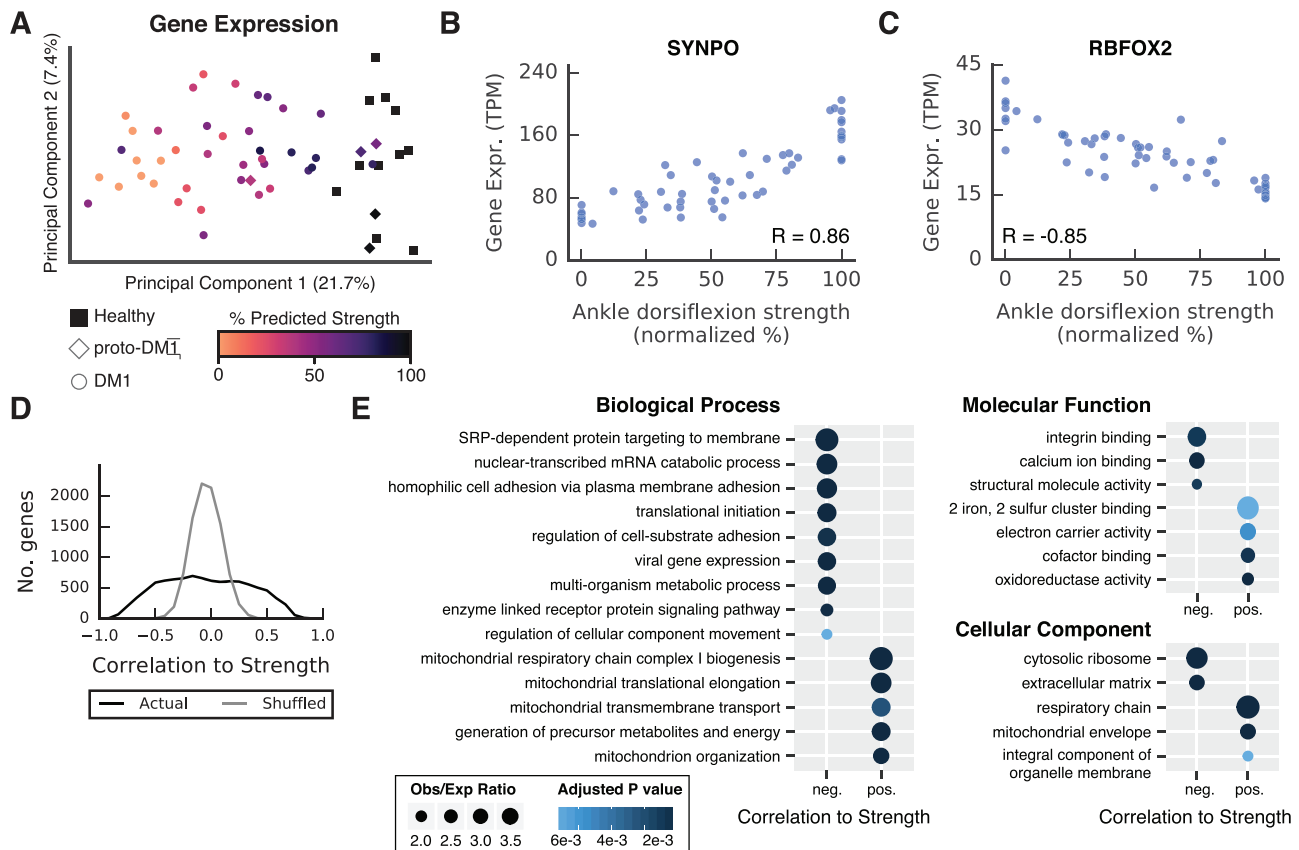
[Supplementary Material, Fig. S3](#)), perhaps reflecting muscle regeneration and the activation of an embryonic splicing program, a phenomenon previously observed in mouse models of muscle injury (26). Future analyses of additional DM2 and DMD samples would be required to assess reproducibility of these patterns.

### Exons in DM1 heart exhibit changes in splicing that may reflect MBNL activity

Several studies of DM1 heart have revealed important exons that may account for symptoms such as cardiac conduction block and cardiac arrhythmia (20,27). However, to date, insufficient data has been available to assess whether splicing changes in

heart also exist along a continuum reflecting underlying MBNL activity or other disease-relevant molecular processes. Here, we estimated  $\Psi$  across 6 healthy hearts and 12 DM1 hearts [including 3 healthy hearts and 3 DM1 hearts from a previously published RNAseq data set (20)]. High-confidence targets were identified by MISO analysis, and  $\Psi$  values were plotted in heatmap form (Fig. 3A). Exons were ranked by the proportion of DM1 hearts showing mis-splicing, and hearts were ranked by total splicing severity (see [Materials and Methods](#)). Some exons showed dysregulation in essentially all hearts (PDLIM3) and others only showed strong dysregulation in a subset of hearts (MBNL1). SCN5A, which contains a mutually exclusive exon pair, exhibits a smooth gradient across DM1 hearts sorted by total splicing dysregulation, suggesting it could be used as a marker of overall mis-splicing. Consistent with previous analyses in





**Figure 4.** Analyses of gene expression changes in DM1 tibialis. (A) Principal component analysis was performed using gene expression estimates across 55 tibialis biopsies. Each point is a single tibialis biopsy from a single individual; squares show healthy individuals, diamonds proto-DM1 individuals and circles DM1 individuals. Points are also colored by normalized ankle dorsiflexion strength. (B) Scatter of gene expression (TPM units) as a function of normalized ankle dorsiflexion strength for SYNPO. The Pearson correlation value is shown. (C) Scatter of gene expression (TPM units) as a function of normalized ankle dorsiflexion strength for RBFOX2. The Pearson correlation value is shown. (D) Histogram of correlation coefficients computed using gene expression relative to normalized ankle dorsiflexion strength (pink). A similar histogram computed using shuffled sample identities is illustrated for comparison (teal). (E) Gene ontology analyses performed using genes that negatively correlate with strength ( $R < -0.5$ ) and those that positively correlate with strength ( $R > 0.5$ ). The size of each point denotes odds ratio and the color denotes the P-value of enrichment.

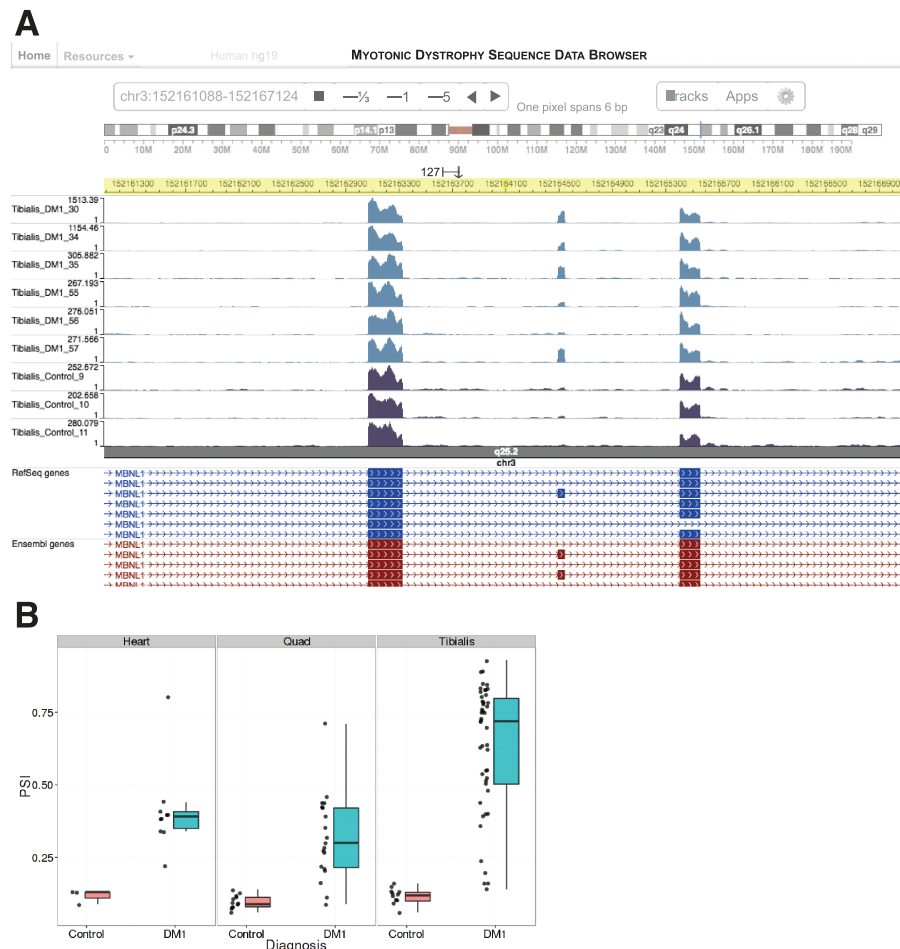
tibialis (22) and data illustrated in Fig. 1,  $\Psi$  values for distinct alternative exons were often highly correlated. For example,  $\Psi$  for SUN1 exon 3 and  $\Psi$  for ABLIM1 exon 11 were both low in healthy heart, but higher in DM1 hearts (Fig. 3B). Correlation coefficients for pairs of all significantly regulated exons were plotted as a histogram and compared to a similar control distribution using shuffled samples (Fig. 3C), revealing a clear signature for concordant mis-splicing changes across the transcriptome.

We also asked whether exons perturbed in DM1 heart overlap with those perturbed in DM1 tibialis. When selecting only exons that were significantly dysregulated in both tissues (50 exons,  $|\Delta\Psi| > 0.1$ ,  $P < 0.01$ ; see Materials and Methods), almost all exons changed in the same direction from healthy to DM1, suggesting that molecular mechanisms of DM pathogenesis are conserved across heart and skeletal muscle (Fig. 3D). Many of these exons have been well validated to be MBNL-dependent targets, including MAPT, NUMA1, MBNL2, NCOR2 and LDB3. However, many exons were also identified to be exclusively regulated in heart (77 exons with  $|\Delta\Psi| > 0.2$  and  $P < 0.01$  in heart but not in muscle; Fig. 3E) or tibialis (108 exons with  $|\Delta\Psi| > 0.2$  and  $P < 0.01$  in tibialis but not heart, Fig. 3F). In some instances, exons dysregulated in one tissue but not the other showed baseline

$\Psi$  values close to zero in the tissue lacking regulation, therefore limiting the possible dynamic range of mis-splicing (Supplementary Material, Fig. S4).

### A gene expression signature for DM1

Gene expression values in tibialis biopsies were estimated by Kallisto (28) (Supplementary Material, Table S5) and subjected to principal component analysis. Healthy samples separated from proto-DM1 and -DM1 samples along the first principal component, and normalized muscle strength also correlated with this component (Fig. 4A). Interestingly, expression of many genes exhibited strong correlations to normalized ankle dorsiflexion strength; SYNPO increases with strength (Fig. 4B) and RBFOX2 decreases with strength (Fig. 4C). Overall, many more genes showed stronger correlations to strength than expected by chance (Fig. 4D), suggesting a signature for gene expression changes in DM1. Separation of genes into those that positively correlated with strength ( $R > 0.5$ ) from those that negatively correlated to strength ( $R < -0.5$ ) showed enrichment of specific Gene Ontology categories (Fig. 4E). For example, genes showing negative correlation, or genes that rise with muscle weakness, showed enrichment in categories such



**Figure 5.** A web-based interface to browse transcriptome data. (A) A screen shot of read density plots (bedgraph format) on the MBNL1 locus. Note the differential inclusion of MBNL1 exon 5, the middle exon visible in the partial gene structure. Several unaffected (control) and DM1 tibialis transcriptomes are loaded. (B) A screen shot obtained from the 'Isoform Browser' page, in which specific exons can be selected. Boxplots and jitter plots displaying  $\Psi$  values for specific exons derived from different tissues are shown; here, the exon selected is MBNL1 exon 5. It is evident that this exon is dysregulated in DM1 heart, quadricep and tibialis tissue.

as 'SRP-dependent protein targeting to membrane', 'integrin binding' and 'extracellular matrix'. Genes that fall with muscle weakness were enriched in categories such as 'mitochondrial respiratory chain complex I biogenesis', 'electron carrier activity' and 'respiratory chain'. These data suggest a myofibroblast response that may be activated in DM1 and a downregulation of core mitochondrial processes characteristic of muscle tissue, possibly due to widespread changes in myonuclear gene expression, changes in cell-type composition of the muscle tissue, or both.

### An online resource for DM1 and related transcriptomes

Although all sequencing data described in this manuscript is deposited at GEO (GSE86356), analyses of gene expression and isoform abundance for multiple samples can require lengthy processing of the data and specific computational expertise/resources. Therefore, we have developed an online resource, [DMseq.org](http://DMseq.org), that makes read density plots across the genome (bedgraph file formats) available to the community (Fig. 5A). The browser is built on the WashU Genome browser, which has been previously used to host extensive volumes of ENCODE RNAseq and CHIPseq data. Because read densities have

been pre-processed, rapid browsing of gene loci is possible by web browser on a computer or mobile device. In addition, we have built additional web pages on the site to allow for query of MISO-computed  $\Psi$  values for alternatively spliced exons for all the human data sets described in this manuscript (Fig. 5B). Finally, additional data sets, such as MBNL1 depletion or CELF overexpression data sets, are also hosted on the same website, can be viewed in the browser and can be cross-compared to the human data sets described here.

### Discussion

Here, we present a comprehensive collection of DM-related transcriptomes. As expected, mis-splicing and gene expression changes clearly stratify disease state and correlate with muscle strength. Although correlations between molecular changes and phenotypes are strong, the precise fraction of toxic CUG RNA that must be depleted or the proportion of MBNL proteins that must be released in order to rescue MBNL-dependent splicing changes and disease phenotypes remains to be precisely defined and will depend on the specific splicing event and phenotype in question. Furthermore, although some mis-spliced exons have been linked to physiological functions and disease symptoms,

the vast majority of molecular changes and their impact on disease progression remain to be characterized. Distinct skeletal muscle groups, for example tibialis anterior and quadriceps, exhibit different levels of transcriptome alterations, and efforts to assess disease progression and develop biomarkers should take account of these differences. Additionally, while some exons exhibit similar changes in DM1 heart and tibialis, many exons are not regulated or expressed at all in one tissue or the other, perhaps due to differential activity of other RNA binding proteins in heart versus tibialis. Some gene expression changes correlate strongly with muscle strength, including some other RNA binding proteins. Interestingly, MBNL1 expression is upregulated in DM1 tibialis biopsies relative to controls, suggesting autoregulatory or compensatory processes. In addition, RBFOX2 expression inversely correlates with muscle strength in DM1 tibialis biopsies, holding certain implications for splicing events co-regulated by MBNL and RBFOX (29,30). Whether these changes occur in muscle fibers or interstitial cells and how they are activated remain topics for future investigation.

In this study, we were not able to perform the same analyses of DM2 that were performed for DM1, because insufficient DM2 samples were available. A key question remains as to whether molecular pathogenesis in DM2 shares many features with DM1; similar efforts in the context of DM2 would help address this question. Similarly, while we analyzed 18 DM1 and unaffected heart transcriptomes, we still could not reliably apply our previously developed quantitative modeling approaches to these samples, because of sample size limitations. Transcriptome analysis of additional heart samples in the future may allow assessment of how cardiac mis-splicing correlates with overall disease severity. The potential of computational modeling to identify biomarkers in these tissues, and also the central nervous system, should motivate future coordinated efforts to obtain tissue samples and generate transcriptomes. In this study, the integrity of RNA derived from biopsy samples was on average higher than that derived from autopsy samples, but some autopsy samples were of the same quality as biopsies. Certainly, estimates of gene expression and splicing from RNA of lower quality may be noisier than those from fully intact RNA, and therefore future efforts utilizing autopsy material should make a best effort to preserve RNA quality.

The [DMseq.org](https://dmseq.org) website has been built to be maximally accessible to investigators who would like to rapidly browse RNAseq read coverage across the genome and access previously computed gene expression and  $\Psi$  values. To facilitate comparisons with transcriptomes derived from mouse models informative about DM1 biology, we have also generated and uploaded genome tracks for a number of published RNAseq transcriptomes. These tracks include transcriptomes from various mouse data sets, e.g. MBNL knockouts (15,31) and CELF overexpressors (32), CLIPseq data sets for MBNL and CELF (15,32) and transcriptomes from postmortem DM1 central nervous system tissues (33).

## Materials and Methods

### Tissue and RNA collection, RNAseq library construction

Tibialis biopsies were obtained as previously described (21). Quadricep biopsies were obtained (23) in accordance with the approval of the local ethics committee (NRESCommittee. EastMidlands-Nottingham2). Informed consent was obtained from all subjects. Autopsy samples were obtained from Baylor

College of Medicine, Stanford University, and University of Minnesota. Tissue samples were homogenized in TRIzol using a handheld tissue homogenizer or pipette followed by Qiagen RNeasy column purification with DNase I treatment. RNAseq libraries were constructed as previously described, using the EpiBio RiboMinus ribosomal depletion kit, followed by dUTP strand-specific RNAseq preparation.

### Read mapping, gene expression quantitation and isoform quantitation

Reads were mapped to the hg19 genome by Hisat2. Gene expression was quantitated using Kallisto, with the hg19\_refseq table as a reference. Isoform  $\Psi$  values were quantitated by MISO using the hg19 v2.0 MISO annotations found here: <https://bio.rc.ufl.edu/pub/ericwang/nonUTRevents.hg19.gff3>. Alternative last exon annotations were generated using UCSC genome transcript annotations and can be found here: <https://bio.rc.ufl.edu/pub/ericwang/ALE.hg19.gff3>.

### Principal Component Analysis (PCA)

For analyses of splicing (Fig. 1B), mapped bam files were subsampled such that the number of mapped reads was ~45 M per sample prior to quantitating isoforms by MISO, because systematic shifts in  $\Psi$  can occur between data sets of widely varying read coverage, especially among lowly expressed exons. PCA analyses were performed using the decomposition.PCA method in the python sklearn package using only splicing events that showed at least 0.4 difference in  $\Psi$  among any two pairs of samples. This threshold focuses the analysis on roughly half of the splicing events with sufficient read coverage across all tibialis samples required to obtain an accurate  $\Psi$  measurement. The first two principal components were plotted.

For analyses of gene expression (Fig. 4A), transcripts per million (TPM) values derived from kallisto were computed (all isoforms for each gene were summed) and only genes with non-zero values across all samples were kept. The logarithm of gene expression was subjected to PCA analysis using the decomposition.PCA method in the python sklearn package. The first two principal components were plotted.

### Sigmoid fit analysis

All  $\Psi$  values for splicing events quantitated in tibialis biopsies were related to [MBNL]<sub>inferred</sub> values (22) using a sigmoid function,  $\Psi = \Psi_{\min} + (\Psi_{\max} - \Psi_{\min}) / (1 + \exp(-\text{slope} * \Psi - \text{EC50}))$ , where each splicing event can have its own values for  $\Psi_{\min}$ ,  $\Psi_{\max}$ , slope and EC50. The 'fit' value was computed as the sum of squares of residual values between observed  $\Psi$  values and the sigmoid curve.

### Motif analysis

For analysis of motifs enriched in and around mis-spliced exons (Fig. 1C), 4-mers were enumerated in the 250 nucleotides upstream and downstream of and within each skipped exon that was significantly dysregulated in DM1 tibialis. The criterion for significant dysregulation was a sigmoid 'fit' of <1 (57 exons aberrantly upregulated and 37 exons aberrantly downregulated). The occurrences of all 4-mers was compared to those obtained



using a set of control exons where the sigmoid 'fit' is  $>1$ . The significance of enrichment for each 4-mer was evaluated using a binomial test, where the expected frequency of each 4-mer in each region was calculated in the control exons.

### Exons regulated in heart

In addition to the 12 samples collected and sequenced here, six additional heart samples from the literature were also included in subsequent analyses (20). MISO  $\Psi$  values were computed as described above. To select and sort heart samples and exons for visualization purposes (Fig. 3A), several filters were used. First, exons strongly dysregulated between unaffected and DM1 hearts were identified using a monotonicity test described previously (32), where  $|\Delta\Psi| > 0.2$  and  $|Z| > 3$ . Then, events were further filtered for those that had a median MISO  $\Psi$  confidence interval  $<0.2$  and standard deviation of  $\Psi < 0.25$  across all unaffected hearts, yielding 39 events. Some of these events were from the same gene, and a single event with the largest  $\Delta\Psi$  between unaffected and DM1 hearts was chosen to visualize in the heatmap, yielding 26 events. The mean  $\Delta\Psi$  across all 26 events was computed ( $\Delta\Psi$  for aberrantly included exons was multiplied by +1 and aberrantly excluded exons by -1) and  $\Psi$  for each event was fit against a sigmoid curve where the mean  $\Delta\Psi$  across all 26 exons was used as the x-axis. Heart samples were sorted by mean  $\Delta\Psi$  and exons were sorted by sigmoid EC50 for the heatmap.

To assess whether exons were regulated in heart and/or tibialis samples (Fig. 3D–F), we filtered for events with  $|\Delta\Psi| > 0.1$  (comparing mean unaffected  $\Psi$  to mean DM1  $\Psi$ ) and  $P < 0.01$  as assessed by ranksum test. For Supplementary Material, Fig. S3, we filtered for exons with  $|\Delta\Psi| > 0.2$  and  $P < 0.01$ .

### Correlations between all splicing events

In Fig. 3C,  $\Psi$  values were correlated between all pairs of exons and shown as a histogram (black line). To obtain a control distribution, the correlations were recomputed using shuffled  $\Psi$  values for one vector of each vector pair. The histogram of the control distribution was shown as a gray line.

## Supplementary Material

Supplementary Material is available at HMG online.

Conflict of Interest statement. None declared.

## Funding

National Institutes of Health (RC2-HG005624 to C.B.B., D.H. and T.C.; R01AR045653 and R01HL045565 to T.C.; NS048843 to C.T.; DP5OD017865 to E.T.W.); Muscular Dystrophy Association (262023 to C.T.); Wellcome Trust (107562/Z/15/Z to D.B.B.); Muscular Dystrophy UK and Myotonic Dystrophy Support Group (D.B.B.).

## References

1. Brook, J.D., McCurrach, M.E., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H., Hunter, K., Stanton, V.P., Thirion, J.P. and Hudson, T. (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**, 799–808.
2. Liquori, C.L., Ricker, K., Moseley, M.L., Jacobsen, J.F., Kress, W., Naylor, S.L., Day, J.W. and Ranum, L.P. (2001) Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science*, **293**, 864–867.
3. Todd, P.K. and Paulson, H.L. (2010) RNA-mediated neurodegeneration in repeat expansion disorders. *Ann. Neurol.*, **67**, 291–300.
4. Kuyumcu-Martinez, N.M., Wang, G.-S. and Cooper, T.A. (2007) Increased steady-state levels of CUGBP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation. *Mol. Cell*, **28**, 68–78.
5. Fernandez-Costa, J.M., Garcia-Lopez, A., Zuñiga, S., Fernandez-Pedrosa, V., Felipe-Benavent, A., Mata, M., Jaka, O., Aiastui, A., Hernandez-Torres, F., Aguado, B. et al. (2013) Expanded CTG repeats trigger miRNA alterations in Drosophila that are conserved in myotonic dystrophy type 1 patients. *Hum. Mol. Genet.*, **22**, 704–716.
6. Rau, F., Freyermuth, F., Fugier, C., Villemain, J.-P., Fischer, M.-C., Jost, B., Dembele, D., Gourdon, G., Nicole, A., Duboc, D. et al. (2011) Misregulation of miR-1 processing is associated with heart defects in myotonic dystrophy. *Nat. Struct. Mol. Biol.*, **18**, 840–845.
7. Kalsotra, A., Singh, R.K., Gurha, P., Ward, A.J., Creighton, C.J. and Cooper, T.A. (2014) The Mef2 transcription network is disrupted in myotonic dystrophy heart tissue, dramatically altering miRNA and mRNA expression. *Cell Rep.*, **6**, 336–345.
8. Jones, K., Wei, C., Iakova, P., Bugiardi, E., Schneider-Gold, C., Meola, G., Woodgett, J., Killian, J., Timchenko, N.A. and Timchenko, L.T. (2012) GSK3 $\beta$  mediates muscle pathology in myotonic dystrophy. *J. Clin. Invest.*, **122**, 4461–4472.
9. Brockhoff, M., Rion, N., Chojnowska, K., Wiktorowicz, T., Eickhorst, C., Erne, B., Frank, S., Angelini, C., Furling, D., Ruegg, M.A. et al. (2017) Targeting deregulated AMPK/mTORC1 pathways improves muscle function in myotonic dystrophy type 1. *J. Clin. Invest.*, **127**, 549–563.
10. Morriss, G.R., Rajapakshe, K., Huang, S., Coarfa, C. and Cooper, T.A. (2018) Mechanisms of skeletal muscle wasting in a mouse model for myotonic dystrophy type 1. *Hum. Mol. Genet.*, **27**, 2789–2804.
11. Zu, T., Gibbens, B., Doty, N.S., Gomes-Pereira, M., Huguet, A., Stone, M.D., Margolis, J., Peterson, M., Markowski, T.W., Ingram, M.A.C. et al. (2011) Non-ATG-initiated translation directed by microsatellite expansions. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 260–265.
12. López Castel, A., Nakamori, M., Tomé, S., Chitayat, D., Gourdon, G., Thornton, C.A. and Pearson, C.E. (2011) Expanded CTG repeat demarcates a boundary for abnormal CpG methylation in myotonic dystrophy patient tissues. *Hum. Mol. Genet.*, **20**, 1–15.
13. Ho, T.H., Charlet, B.N., Poulos, M.G., Singh, G., Swanson, M.S. and Cooper, T.A. (2004) Muscleblind proteins regulate alternative splicing. *EMBO J.*, **23**, 3103–3112.
14. Batra, R., Charizanis, K., Manchanda, M., Mohan, A., Li, M., Finn, D.J., Goodwin, M., Zhang, C., Sobczak, K., Thornton, C.A. and Swanson, M.S. (2014) Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease. *Mol. Cell*, **56**, 311–322.
15. Wang, E.T., Cody, N.A., Jog, S., Biancolella, M., Wang, T.T., Treacy, D.J., Luo, S., Schroth, G.P., Housman, D.E., Reddy, S. et al. (2012) Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell*, **150**, 710–724.

16. Charlet, B.N., Savkur, R.S., Singh, G., Philips, A.V., Grice, E.A. and Cooper, T.A. (2002) Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. *Mol. Cell*, **10**, 45–53.
17. Mankodi, A., Takahashi, M.P., Jiang, H., Beck, C.L., Bowers, W.J., Moxley, R.T., Cannon, S.C. and Thornton, C.A. (2002) Expanded CUG repeats trigger aberrant splicing of ClC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy. *Mol. Cell*, **10**, 35–44.
18. Fugier, C., Klein, A.F., Hammer, C., Vassilopoulos, S., Ivarsson, Y., Toussaint, A., Tosch, V., Vignaud, A., Ferry, A., Messaddeq, N. et al. (2011) Misregulated alternative splicing of BIN1 is associated with T tubule alterations and muscle weakness in myotonic dystrophy. *Nat. Med.*, **17**, 720–725.
19. Tang, Z.Z., Yarotsky, V., Wei, L., Sobczak, K., Nakamori, M., Eichinger, K., Moxley, R.T., Dirksen, R.T. and Thornton, C.A. (2012) Muscle weakness in myotonic dystrophy associated with misregulated splicing and altered gating of Ca(V)1.1 calcium channel. *Hum. Mol. Genet.*, **21**, 1312–1324.
20. Freyermuth, F., Rau, F., Kokunai, Y., Linke, T., Sellier, C., Nakamori, M., Kino, Y., Arandel, L., Jollet, A., Thibault, C. et al. (2016) Splicing misregulation of SCN5A contributes to cardiac-conduction delay and heart arrhythmia in myotonic dystrophy. *Nat. Commun.*, **7**, 11067.
21. Nakamori, M., Sobczak, K., Puwanant, A., Welle, S., Eichinger, K., Pandya, S., Dekdebrun, J., Heatwole, C.R., McDermott, M.P., Chen, T. et al. (2013) Splicing biomarkers of disease severity in myotonic dystrophy. *Ann. Neurol.*, **74**, 862–872.
22. Wagner, S.D., Struck, A.J., Gupta, R., Farnsworth, D.R., Mahady, A.E., Eichinger, K., Thornton, C.A., Wang, E.T. and Berglund, J.A. (2016) Dose-dependent regulation of alternative splicing by MBNL proteins reveals biomarkers for myotonic dystrophy. *PLoS Genet.*, **12**, e1006316.
23. Wojciechowska, M., Sobczak, K., Kozłowski, P., Sedehizadeh, S., Wojtkowiak-Szlachcic, A., Czubak, K., Markus, R., Lusakowska, A., Kaminska, A. and Brook, J.D. (2018) Quantitative methods to monitor RNA biomarkers in myotonic dystrophy. *Sci. Rep.*, **8**, 5885.
24. Katz, Y., Wang, E.T., Airolidi, E.M. and Burge, C.B. (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods*, **7**, 1009–1015.
25. Goers, E.S., Purcell, J., Voelker, R.B., Gates, D.P. and Berglund, J.A. (2010) MBNL1 binds GC motifs embedded in pyrimidines to regulate alternative splicing. *Nucleic Acids Res.*, **38**, 2467–2484.
26. Orengo, J.P., Ward, A.J. and Cooper, T.A. (2011) Alternative splicing dysregulation secondary to skeletal muscle regeneration. *Ann. Neurol.*, **69**, 681–690.
27. Philips, A.V., Timchenko, L.T. and Cooper, T.A. (1998) Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science*, **280**, 737–741.
28. Bray, N.L., Pimentel, H., Melsted, P. and Pachter, L. (2016) Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.*, **34**, 525–527.
29. Klinck, R., Fourrier, A., Thibault, P., Toutant, J., Durand, M., Lapointe, E., Caillet-Boudin, M.-L., Sergeant, N., Gourdon, G., Meola, G. et al. (2014) RBFOX1 cooperates with MBNL1 to control splicing in muscle, including events altered in myotonic dystrophy type 1. *PLoS One*, **9**, e107324.
30. Venables, J.P., Lapasset, L., Gadea, G., Fort, P., Klinck, R., Irimia, M., Vignal, E., Thibault, P., Prinos, P., Chabot, B. et al. (2013) MBNL1 and RBFOX2 cooperate to establish a splicing programme involved in pluripotent stem cell differentiation. *Nat. Commun.*, **4**, 2480.
31. Charizanis, K., Lee, K.-Y., Batra, R., Goodwin, M., Zhang, C., Yuan, Y., Shiue, L., Cline, M., Scotti, M.M., Xia, G. et al. (2012) Muscleblind-like 2-mediated alternative splicing in the developing brain and dysregulation in myotonic dystrophy. *Neuron*, **75**, 437–450.
32. Wang, E.T., Ward, A.J., Cherone, J.M., Giudice, J., Wang, T.T., Treacy, D.J., Lambert, N.J., Freese, P., Saxena, T., Cooper, T.A. et al. (2015) Antagonistic regulation of mRNA expression and splicing by CELF and MBNL proteins. *Genome Res.*, **25**, 858–871.
33. Goodwin, M., Mohan, A., Batra, R., Lee, K.-Y., Charizanis, K., Fernández Gómez, F.J., Eddarkaoui, S., Sergeant, N., Buée, L., Kimura, T. et al. (2015) MBNL sequestration by toxic RNAs and RNA misprocessing in the myotonic dystrophy brain. *Cell Rep.*, **12**, 1159–1168.