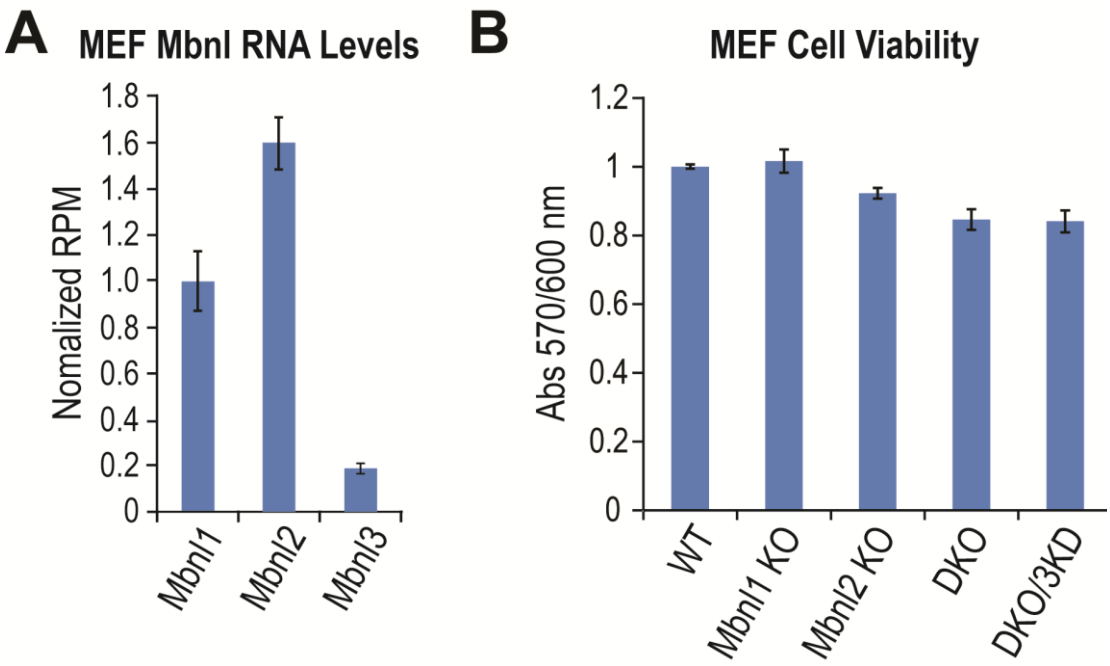

Figure S1. Mbnl Knockout MEF RNA Levels and Cell Viability
(A) Transcript abundance in WT primary MEFs was assessed using total reads per million (RPM) mapped to *Mbnl* polyA sites. RPM data were normalized to Mbnl1 levels.
(B) Cell viability of WT, Mbnl1 KO, Mbnl2 KO, DKO and DKO/3KD MEFs. Related to Figure 1.

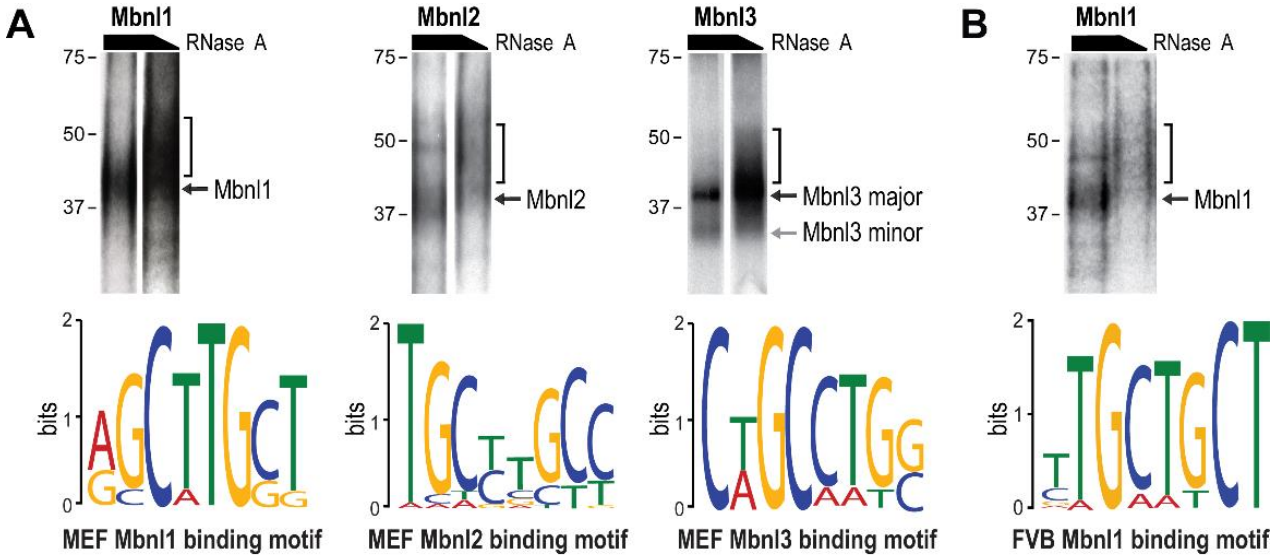

Figure S2. Selection of RNA-Mbnl Complexes for HITS-CLIP

(A) Gel analysis of MEF Mbnl-protein complexes followed UV-light induced crosslinking and treatment with high (thick black bar) and low (sloping black bar) RNase A concentrations. The RNA-protein complexes isolated for HITS are indicated by a bracket and both major (38 kDa) and minor (27 kDa) Mbnl3 isoforms are indicated. Although the 27 kDa minor isoform of Mbnl3 (Poulos et al., 2013) was also detectably crosslinked to MEF RNA, RNA-protein complexes for the larger isoform were primarily selected to be consistent with the Mbnl1 and Mbnl2 datasets. Preferred Mbnl binding sites motifs (R/YGCY) are illustrated below each CLIP gel and contain either a single (Mbnl3) or multiple (Mbnl1, Mbnl2) R/YGCY motifs. Related to Figure 2.

(B) As (A) but Mbnl1 binding to wild type (FVB) mouse quadriceps muscle RNA. Related to Figure 2.

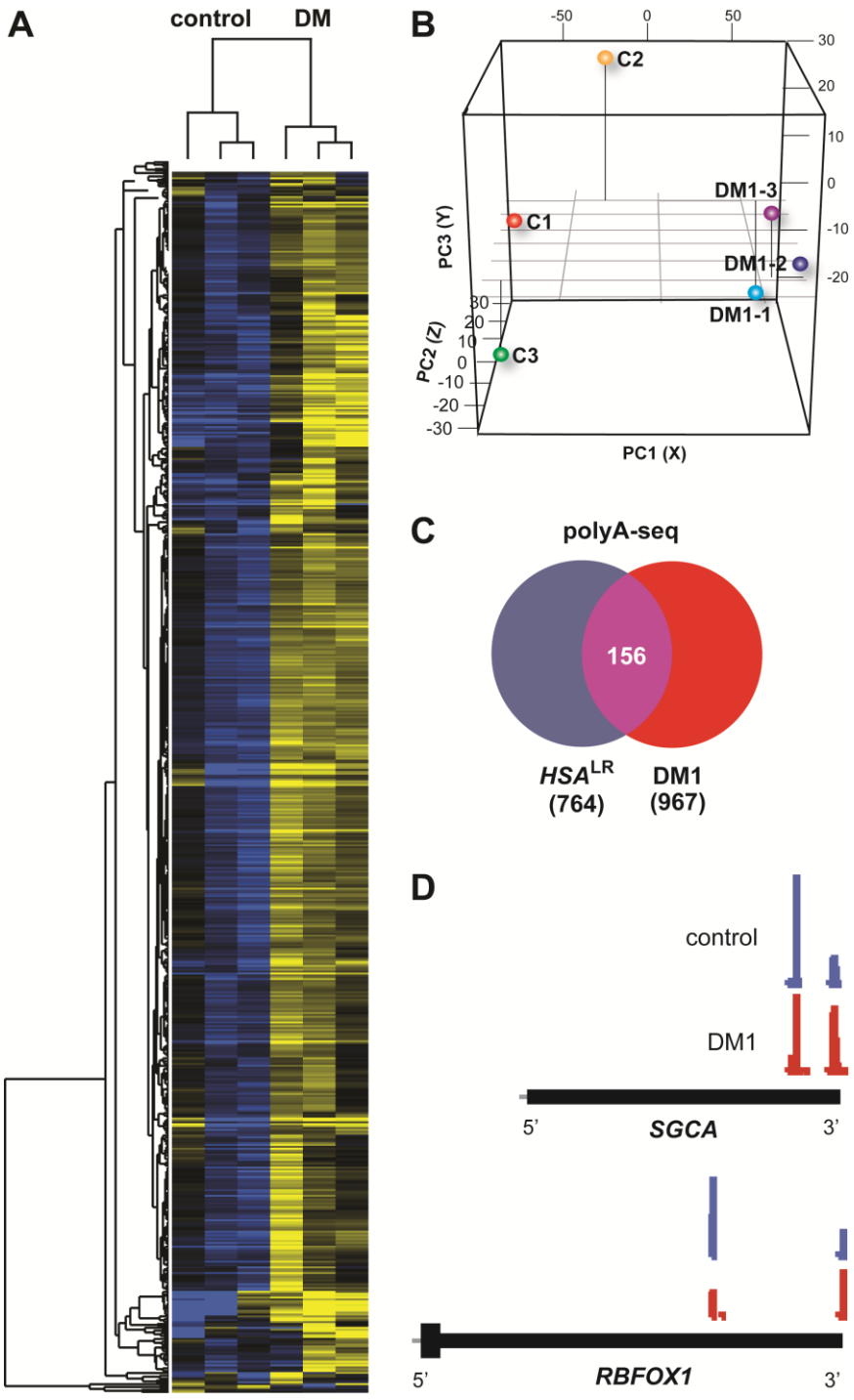

Figure S3. Altered APA in Human DM1 Muscle

(A) Hierarchical clustering of $\log_2(n+1)$ transformed data of polyA sites (top 600) in control and DM1 human autopsy muscle where n is the read count of each polyA site.

(B) Principal component analysis (PCA) of $\log_2(n+1)$ transformed data of polyA sites (top 2500) for control and DM1 human autopsy muscle.

(C) Venn diagram of PolyA-seq showing overlap between the mouse *HSA*^{LR} and human DM1 data sets.

(D) Wiggle plots of PolyA-seq results for SGCA and RBFOX1. Related to Figure 6.

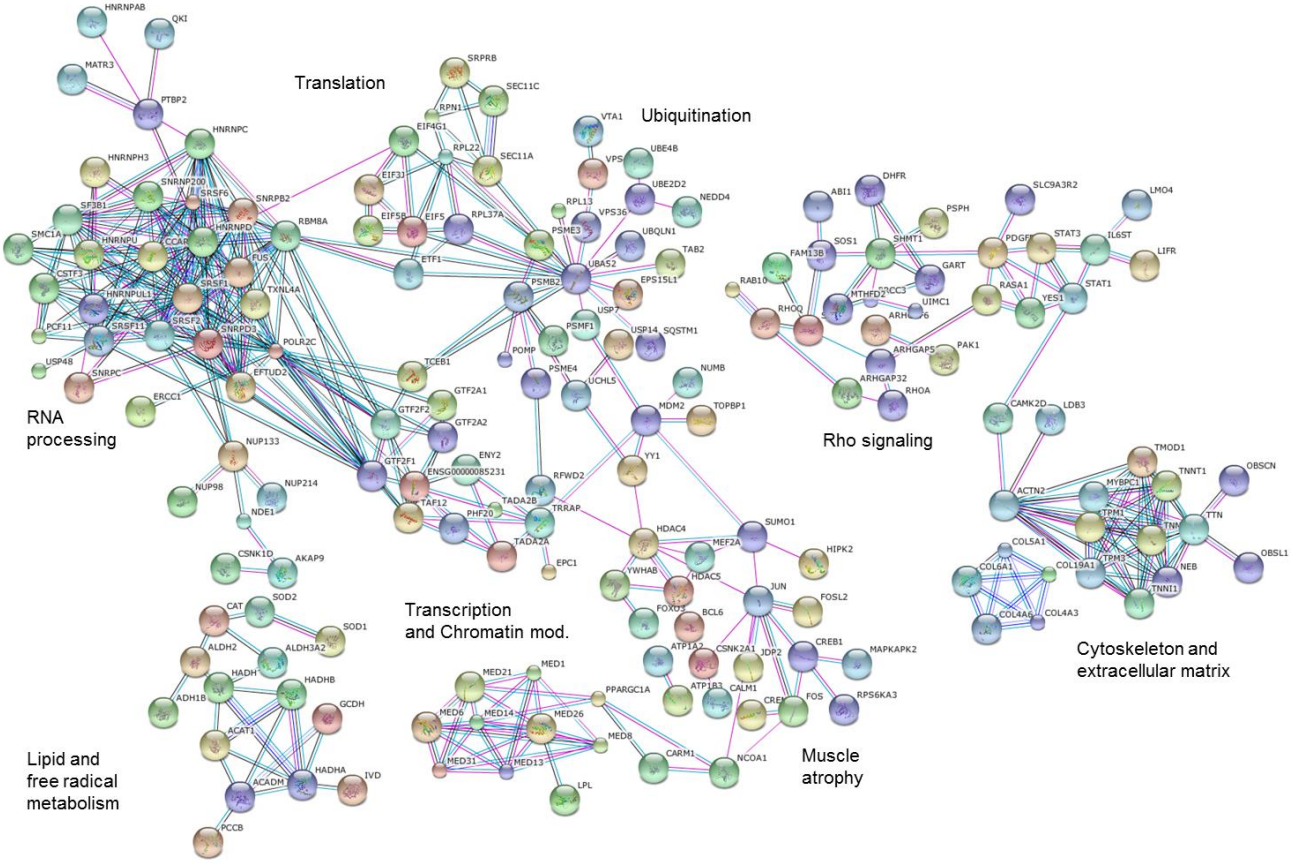

Figure S4. Misregulated APA gene networks in DM1 muscle

STRING v9.1 (Franceschini et al., 2013) was used to evaluate gene networks affected at the APA level in DM1 skeletal muscle. Cellular pathways for the regulation of RNA processing, ubiquitination, muscle contraction and cytoskeleton, transcription, lipid metabolism and muscle maintenance are affected. Related to Figure 6.

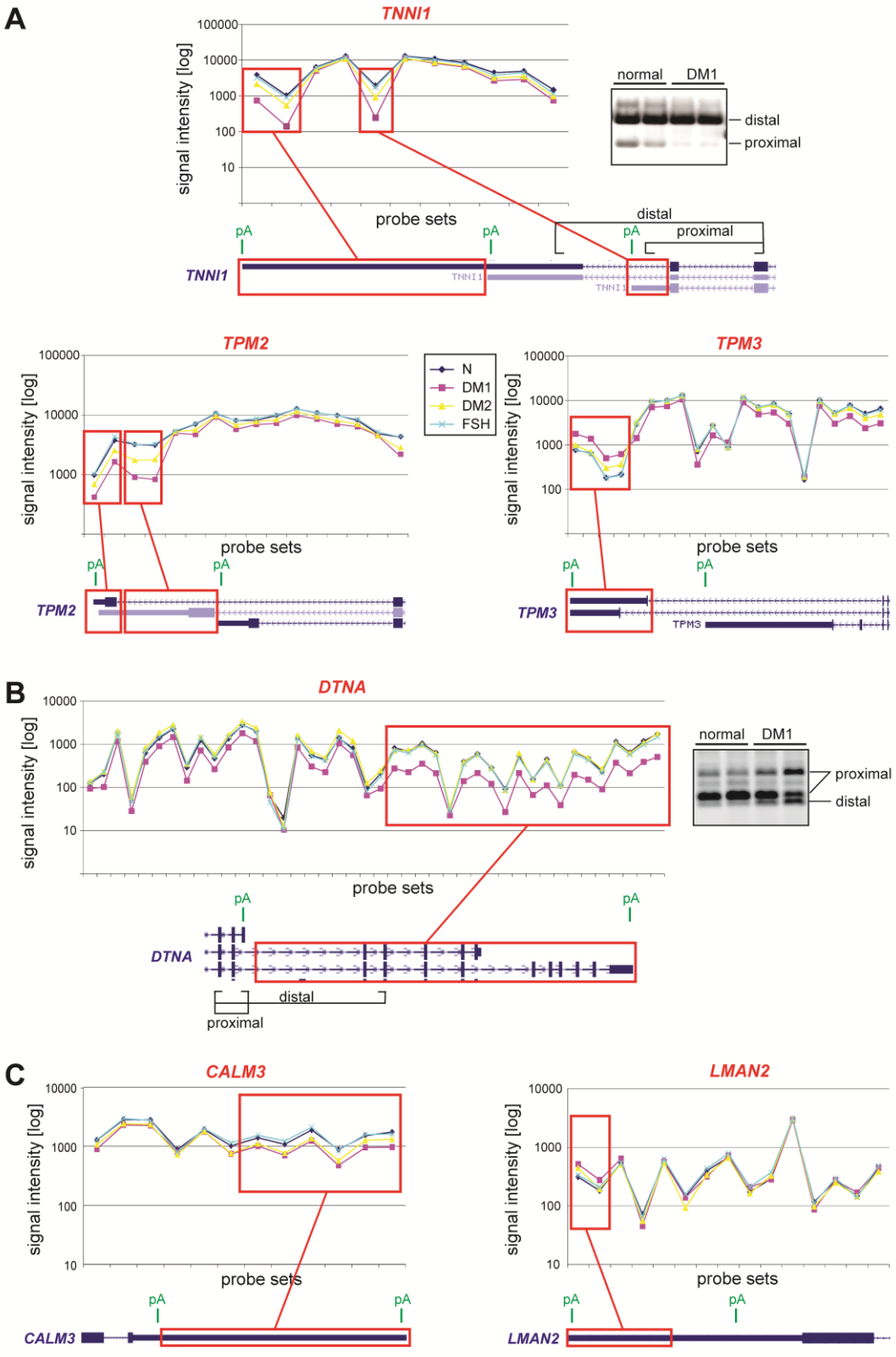

Figure S5. DM-specific APAs based on AllExon array analysis

(A) Three examples (TNNI1, TPM2, TPM3) of alternative 3' end formation where 3' UTR sequence and length are mainly affected.

(B) Example (DTNA) of APA shift that results in a much longer protein coding region.

(C) APA examples (CALM3, LMAN2) selected based on polyA-seq with corresponding changes in AllExon analysis that are statistically significant but failed the restrictive initial selection criteria so are not listed in Table S5.

Related to Figure 7.

Table S1. PolyA-seq Analysis (Related to Figures 2, 4, 5, 6)

(A) MEF WT vs *Mbnl1*$^{-/-}$; *Mbnl2*$^{-/-}$ (DKO)

(B) MEF WT vs DKO/Mbnl3 knockdown (DKO/3KD)

(C) WT adult vs WT postnatal day 1 (P1) quadriceps muscle

(D) WT adult vs *HSA*^{LR} adult muscle

(E) Control vs DM1 muscle

(F) pA-supported Mbnl HITS-CLIP tags (>20 tags/cluster)

(G) pA shifts only in DKO/3KD MEFs

(H) pA shifts in which DKO/3KD > DKO MEFs

Table S2. HITS-CLIP Analysis (Related to Figures 2, 5)

(A) Mbnl1 WT MEF

(B) Mbnl2 WT MEF

(C) Mbnl3 DKO MEF

(D) Mbnl3 WT MEF

(E) Mbnl1 FVB WT muscle

(F) Cpsf6 WT MEF

(G) Cpsf6 DKO MEF

(H) HITS-CLIP summary

(I) Noncoding RNAs: Mir (note Mir1a-1 in Mbnl1 HITS-CLIP of quadriceps)

(J) Noncoding RNAs: snoRNA

Table S3. PCR Primers for PolyA-seq Validation and Minigene Analysis (Related to Figures 1, 3)

Table S4. Human Skeletal Muscle Autopsy Samples (Related to Figure 6)

Table S5. Candidate APA Events Identified by AllExon Microarrays (Related to Figure 7)

Table S6. Human and Mouse Primers for AllExon Array Studies (Related to Figure 7).

Table S7. RT-PCR Verified APAs Selected from AllExon Microarray Dataset

Gene	*Developmental APA switch in mouse by RT-PCR analysis*	*PolyA-seq (FDR values)*			
		N vs DM1	**WT vs DKO/3KD**	**WT vs P1**	**WT vs *HSA*^{LR}**
Confirmed DM1 specificity and MBNL dependence					
TPM1	ND	0,025419	0	3,12E-200	0,002068
DTNA	yes	4,06E-11	1,87E-07	6,40E-34	7,94E-05
PDLIM5	yes	4,74E-11	3,53E-18	1,20E-30	0,000721
TPM3	yes	1,71E-11	9,45E-06	0,017647	-
DNAJB6	yes	-	0	1,06E-08	-
Confirmed DM1 specificity but not confirmed MBNL dependence					
TNNI1	ND	2,25E-12	-	3,81E-12	-
PDLIM3	ND	1,13E-42	-	-	-
SPTB	yes	1,31E-09	-	-	-
LDB3	yes	8,81E-76	-	1,03E-52	0,028083
SETD3	ND	5,00E-06	-	-	-
Not confirmed DM1 specificity but confirmed MBNL dependence					
ASPH	yes	-	4,92E-09	3,04E-06	-
LMNA	no	-	0	-	-
CELF1	yes	-	2,54E-06	4,77E-06	-
Not confirmed DM1 specificity and MBNL dependence					
ARHGEF7	no	-	-	-	-
PLIN2	yes	-	-	-	-
TGFB1	ND	-	-	-	-
CIRBP	yes	-	-	-	-

ND, not determined; Related to Figure 7.

Supplemental Experimental Procedures

Protein Analysis

Cell lysates were prepared, and subcellular localization was determined, following cell fractionation as described previously (Poulos et al., 2013). Proteins were detected by immunoblotting using rabbit polyclonal antibody (rpAb) anti-Mbnl1 A2764 (gift of C. Thornton, University of Rochester), mouse monoclonal (mAb) anti-Mbnl2 3B4 (Santa Cruz Biotechnology), anti-Gapdh mAb 6C5 (Abcam), rpAb anti-Ldha (Cell Signaling Technology) and rpAb anti-Mbnl3 Mb3/7G (Poulos et al, 2013).

RNA-FISH

For FISH, 10 μm cryosections were prepared from WT FVB and *HSA*^{LR} mice. Sections were air dried for 30 min, fixed in 2% formalin at 4°C for 30 min, incubated in 2% acetone/PBS at -20°C, pre-hybridized for 30 min at 42°C (30% formamide/2x SSC) and hybridized for 2 hr at 42°C with CAG_{10} DNA probe (30% formamide, 2x SSC, 0.02% BSA, 67 ng/μl yeast tRNA, 2 mM vanadyl adenosine, 1 ng/μl Cy3-CAG_{10}). After hybridization, the sections were washed in 30% formamide/2xSSC at 45°C for 30 min and incubated with primary antibodies overnight at 4°C (1:1000 anti-Mbnl1 polyclonal antibody A2764, 1:100 anti-Mbnl2 monoclonal 3B4, 1:1000 anti-Mbnl3 polyclonal Mb3/7G). Confocal microscopy was performed using a Leica SP-5 microscope.

HITS-CLIP

Quadriceps muscles from WT FVB mice (3 months, n = 3 each genotype) were dissected, frozen in liquid nitrogen, ground to a fine powder and crosslinked with UV-light (Stratalinker 1800, Stratagene). For MEFs, 10 cm plates (n = 2 each genotype), were UV-crosslinked followed by scraping cells in ice-cold PBS and centrifugation. For immunoprecipitation, protein lysates were denatured by adding SDS (1% final concentration), heating (100°C/5 min) and subsequent dilution to a final SDS concentration of 0.1% in PXL lysis buffer (Chi et al., 2009) prior to the addition of anti-Mbnl1 (rpAb A2764), Mbnl2 (mAb 3B4) and Mbnl3 (rpAb Mb3/7G) antibodies (5 μg each). RNA tags were generated as described (Charizanis et al., 2012) using RNase A (concentrations of 55 U/ml and 0.55 U/ml for high and low RNase, respectively) and cDNA libraries were generated using RNA linkers and primers as described (Chi et al., 2009).

Raw reads obtained from the Illumina pipeline included a 4 nt barcode (3 random positions and a G at the 4th position) at the 5' end followed by the CLIP tag and the 3' adaptor sequence. Results were filtered to remove low quality reads by requiring a minimum score of 20 in barcode positions and an average score of 20 in the following 25 nucleotides. The filtered reads were then aligned to the mouse reference genome (mm9) by Novoalign (http://www.novocraft.com) allowing iterative trimming at the 3'-end (Zhang and Darnell, 2011). Potential PCR duplicates, as judged from the starting position of genomic mapping and the barcode sequences, were removed to identify unique CLIP tags as described (Charizanis et al., 2012). Crosslink-induced mutation site (CIMS) analysis was performed as described (Zhang and Darnell, 2011) and deletions were identified using Galaxy software. For motif enrichment, de novo motif searches were accomplished using the MEME suite and randomly selected sequences (n = 2000) around CIMS (from −7 to +7 nt). Motif enrichment for YGCY, YCGY and YCAY sequences was also performed around Mbnl CLIP clusters (± 500 nt, cluster height > 5) for FVB WT quadriceps and MEFs. Mbnl binding motifs were determined as described using the top 500 CLIP peaks from MEF Mbnl1, Mbnl2 and Mbnl3 HITS-CLIP datasets (Poulos et al., 2013). Gene ontology (GO) enrichment analysis was performed using Ingenuity Pathway Analysis (IPA Ingenuity Systems, www.ingenuity.com). Canonical signaling pathways were listed according to GO enrichment as –log (p-value) with $p < 0.001$. Composite RNA maps were generated after mapping and calculating the CLIP read density 200 bp upstream and downstream of skipped and utilized PAs (*P* value <0.00005). Polyadenylation sites were divided into included and skipped categories due to increase or decrease in PolyA-seq read density at that particular PA in WT versus DKO MEFs. The number of unique CLIP tags at each cleavage site were counted for each included and skipped PA and a density chart generated.

PolyA-seq Data

The raw sequencing reads were mapped to the reference genome (mm10) using the program OLego (Wu and Zhang, in preparation) with a seed size of 13 nt and allowing ≤ two mismatches or insertions or deletions. Since the antisense strand is sequenced, mapped reads were reverse complemented to obtain the sense strand sequence, and the most 3' end position was recorded since this position corresponds to the last nucleotide preceding the polyA tail (pA site).

To define robust polyA sites, reads from all samples were pooled and supporting reads for each candidate polyA site were counted. Candidate polyA sites that likely resulted from internal priming from A-rich RNA sequences were then removed using a log-odds model established in a previous study (Derti et al., 2012). These cleaned sites were further filtered to remove those supported by few reads (<10 reads) and then grouped so that sites in each cluster are ≤ 25 nt with each other (Tian et al., 2005). In each cluster, the site supported by the largest number of reads was selected as a representative, denoted unique pA site. Finally, unique polyA sites was assigned to overlapping RefSeq and UCSC annotated genes. If a polyA site did not overlap with a known gene, but overlapped with the extended 3' UTR (1 kb extension) of a transcript, it was assigned to that gene. To identify differential polyA site usage, we considered each pair of polyA sites assigned to the same gene using the number of supporting reads in each group of samples. Statistical analysis was performed using Fisher's exact test, followed by Benjamini correction (Benjamini and Hochbergh, 1995) for multiple tests and calculation of proportional change of polyA site usage. To minimize multiple test correction, only those relatively abundant polyA sites were considered (coverage≥20, or {site1_WT+site2_WT≥20, site1_KO+site2_KO≥20, site1_WT+site1_KO≥20, site2_WT+site2_KO≥20}). For hierarchical clustering and principal component analysis, counts per million (CPM) normalized data was $\log_2(n+1)$ transformed, filtered to exclude low read number ($<\log_2 5$), centered around the mean, and clustered with Cluster 3.0 or subjected to PCA analysis using R. Hierarchical clustering was visualized with Java Tree view and PCA analysis was plotted with the gplot function and rgl library in R.

RNA Immunoprecipitation Protocol (RIP)

RIP assays were performed as described previously (Niranjanakumari et al., 2002) using WT and DKO MEFs. Cells were treated with 0.1% formaldehyde for 1 hr followed by RNA fragmentation (0.1 U/µl RNase T1), immunoprecipitation with antibodies against Cpsf-160 (Bethyl), Cstf2 (gift of C. MacDonald) and Tardbp (Proteintech), heat-induced crosslinking reversal and then RT-PCR using primers specific for short (pA1) and long (pA2) 3'UTRs.

Microarray Analysis

Human AllExon array analysis was performed as described previously for eight samples from DM1, seven samples from DM2, eight samples from FSHD patients (disease control) and eight samples from non-dystrophic individuals (normal control) (Nakamori et al., 2013). Briefly, total cellular RNA was isolated from muscle tissue using Trizol (Invitrogen). End-labeled sense-strand cDNA fragments were generated using GeneChip WT Sense Target Labeling kits (Affymetrix) and hybridized to Human Exon 1.0 ST arrays (Affymetrix) according to the manufacturer's recommendations and hybridization signal intensity was determined by laser scanning. Fluorescence intensity data were extracted and normalized using the quartile sketch algorithm of Exon Array Computational Tool software (ExACT, Affymetrix). The normalized signal from each probe for a single exon (typically 4 probes) was then condensed to a single value representing a probe set using PLIER (Probe Logarithmic Intensity Error), which generates a signal estimate using a model fit (these values are plotted on graphs for all probe sets representing single gene). Some exons, especially long 3' exons, are represented by more than one probe set. To remove probes with a high likelihood of false detection additional filtering criteria were applied, which resulted in the generation of a list of candidate exons showing differential expression in DM (see Nakamori et al. 2013 for more details). To prioritize candidates, probe sets (exons) were selected with a >2-fold change in either DM1 or DM2 vs. normal controls and *P* values <0.01 for the DM1 vs. normal comparison and *P* values <0.05 in the DM2 vs. normal comparison. All probe sets showing significant signal changes in FSHD samples were excluded. The group of potential APA isoforms (Table S5) was chosen following visual inspection of each candidate on graphs showing signal intensity for all probe sets representing single genes for four groups of samples and localization of differentially expressed exons on RefSeq gene structures (Figures 7 and S5).

Effects of DM1 and DM2 on 3' end formation revealed by microarray analysis were validated by RT-PCR employing 3 primers including one forward primer and one reverse primer in each of the two (or three) alternative isoforms. cDNAs were generated using oligo-dT and random hexamer primers. PCR primer sets for human and mouse APA events are listed in Table S6. PCR products were separated on agarose gels, stained with SybrGreen, and analyzed on a laser scanner (Typhoon, GE Healthcare Life Sciences) as previously described (Nakamori et al., 2013). Candidate confirmation was performed using replicate assays comparing postmortem DM1 vs. normal muscle RNA (n = 2-5

DM1 samples vs. 2 vastus lateralis samples from healthy controls). Positively confirmed APA events are listed in Table S7. For this group of events, RT-PCR assays were also performed on mouse quadriceps RNA from wild-type mice at embryonic day 18 (E18), postnatal day 2 (P2), P8, P20 and adult (8 weeks of age) mice.

Supplemental References

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B Stat. Methodol. *57*, 289-300.

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.

Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature *460*, 479-486.

Derti, A., Garrett-Engele, P., Macisaac, K.D., Stevens, R.C., Sriram, S., Chen, R., Rohl, C.A., Johnson, J.M., and Babak, T. (2012). A quantitative atlas of polyadenylation in five mammals. Genome Res. *22*, 1173-1183.

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C.*, et al.* (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucl. Acids Res. *41*, D808-815.

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.

Niranjanakumari, S., Lasda, E., Brazas, R., and Garcia-Blanco, M. (2002). Reversible cross-linking combined with immunoprecipitation to study RNA-protein interactions in vivo. Methods *26*, 182-190.

Poulos, M.G., Batra, R., Li, M., Yuan, Y., Zhang, C., Darnell, R.B., and Swanson, M.S. (2013). Progressive impairment of muscle regeneration in muscleblind-like 3 isoform knockout mice. Hum. Mol. Genet. *22*, 3547-3558.

Tian, B., Hu, J., Zhang, H., and Lutz, C.S. (2005). A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucl. Acids Res. *33*, 201-212.

Zhang, C., and Darnell, R.B. (2011). Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nature Biotech. *29*, 607-614.