nature structural & molecular biology

Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy

Hongqing Du[1], Melissa S Cline[1], Robert J Osborne[2], Daniel L Tuttle[3], Tyson A Clark[4], John Paul Donohue[1], Megan P Hall[1], Lily Shiue[1], Maurice S Swanson[3], Charles A Thornton[2] & Manuel Ares Jr[1]

The common form of myotonic dystrophy (DM1) is associated with the expression of expanded CTG DNA repeats as RNA (CUG^{exp} RNA). To test whether CUG^{exp} RNA creates a global splicing defect, we compared the skeletal muscle of two mouse models of DM1, one expressing a CTG^{exp} transgene and another homozygous for a defective muscleblind 1 (*Mbnl1*) gene. Strong correlation in splicing changes for ~100 new Mbnl1-regulated exons indicates that loss of Mbnl1 explains >80% of the splicing pathology due to CUG^{exp} RNA. In contrast, only about half of mRNA-level changes can be attributed to loss of Mbnl1, indicating that CUG^{exp} RNA has Mbnl1-independent effects, particularly on mRNAs for extracellular matrix proteins. We propose that CUG^{exp} RNA causes two separate effects: loss of Mbnl1 function (disrupting splicing) and loss of another function that disrupts extracellular matrix mRNA regulation, possibly mediated by Mbnl2. These findings reveal unanticipated similarities between DM1 and other muscular dystrophies.

Myotonic dystrophy (DM) is a triplet-repeat expansion disease, one of a group that includes Huntington's disease, Fragile X syndrome and Friedreich's ataxia[1,2]. The common form, DM1, is caused by a CTG-repeat expansion (CTG^{exp}) in the 3′ untranslated region (UTR) of the myotonic dystrophy protein kinase (*DMPK*) gene, leading to myotonia, muscle degeneration, reduced heart function, ocular cataracts and nervous system dysfunction[3]. RNA containing CUG repeats (CUG^{exp} RNA) accumulates in nuclear foci[4,5]. Based on its autosomal dominant inheritance, a leading hypothesis for the cause of DM1 is that CUG^{exp} RNA is toxic. Mice engineered to express CUG^{exp} RNA show many symptoms of myotonic dystrophy[6,7]. In addition to altering RNA splicing (see below), CUG^{exp} RNA has been proposed to disrupt a wide variety of cellular processes through several mechanisms, including "leaching" of transcription factors[8], being processed into small RNAs that trigger inappropriate gene silencing[9] or activating PKC-dependent signaling pathways[10]. Among these, the contribution of splicing perturbations is most clear given the physiological relevance of several splicing alterations that occur when CUG^{exp} RNA is expressed[5,7,11]. For example, aberrant splicing of transcripts from the chloride channel gene *Clcn1* is responsible for the myotonia in skeletal muscle[12–15], and aberrant splicing of insulin receptor transcripts is associated with insulin resistance[16].

Among proteins that bind CUG^{exp} RNA are homologs of *Drosophila melanogaster muscleblind* (*muscleblind*-like proteins Mbnl1–3)[17,18]. Mbnl proteins contain four CCCH-type zinc fingers that recognize a YGCY motif repeated within the CUG repeat (CUGCUG) and can function in splicing regulation[19–21]. CUG repeats can form long double-stranded RNA structures *in vitro*[22]; however, crystal structures of Mbnl1 zinc fingers in complex with RNA indicate that the Watson-Crick faces of the binding-site 5′-GC-3′ dinucleotide residues are buried in the protein and unavailable for duplex formation[20]. The colocalization of CUG^{exp} RNA and Mbnl1 in nuclear foci suggests a place where sequestration occurs[4,11,17].

Beside Mbnl1, the altered function of splicing factors CUGBP1 and hnRNP H have been proposed to serve a role in DM1 pathogenesis[23–26]. To determine the role of Mbnl1, a homozygous Mbnl1 mutant mouse (*Mbnl1*$^{\Delta E3/\Delta E3}$) was created[27]. Like mice expressing CUG^{exp} RNA, mice deficient in Mbnl1 show characteristics of myotonic dystrophy including aberrant splicing[11,19,27]. Some but not all of the DM-like symptoms and aberrant splicing of six exons are rescued by adeno-associated virus (AAV)-mediated expression of Mbnl1 in CUG^{exp} RNA–expressing mice[28], leaving open the possibility that CUG^{exp} RNA has other mechanisms of action. In addition, the broader impact on splicing of the loss or sequestration of Mbnl1 is unknown beyond the few genes tested so far.

To determine the extent to which the loss of Mbnl1 explains the splicing and gene-expression defects caused by CUG^{exp} RNA, we compared mRNA in skeletal muscle of *HSA*^{LR} mice[7] expressing CUG^{exp} RNA to those in *Mbnl1*$^{\Delta E3/\Delta E3}$ mice[27] using splicing-sensitive microarrays[29,30]. We found that global splicing perturbations were remarkably congruent in these mice and identified more than 200 splicing events altered upon loss of Mbnl1. Testing of human exons orthologous to these new mouse Mbnl1-dependent exons suggests that humans with DM1 suffer many of the same mis-splicing events, identifying new potential splicing markers for the human disease. Mbnl1 RNA binding sequences

[1]RNA Center, Department of Molecular, Cell and Developmental Biology, Sinsheimer Labs, University of California, Santa Cruz, California, USA. [2]Neuromuscular Disease Center, Department of Neurology, University of Rochester School of Medicine and Dentistry, Rochester, New York, USA. [3]Department of Molecular Genetics & Microbiology, University of Florida, College of Medicine, Gainesville, Florida, USA. [4]Affymetrix Inc. Santa Clara, California, USA. Correspondence should be addressed to M.A. (ares@biology.ucsc.edu).

Received 16 July 2009; accepted 14 October 2009; published online 24 January 2010; doi:10.1038/nsmb.1720

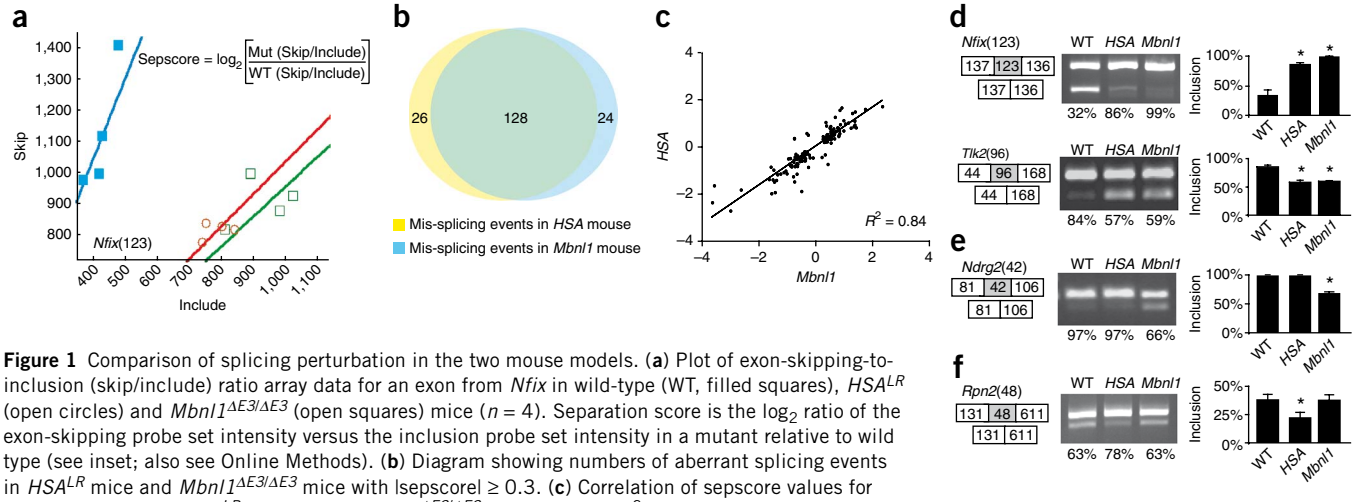

Figure 1 Comparison of splicing perturbation in the two mouse models. (**a**) Plot of exon-skipping-to-inclusion (skip/include) ratio array data for an exon from *Nfix* in wild-type (WT, filled squares), HSA^{LR} (open circles) and $Mbnl1^{\Delta E3/\Delta E3}$ (open squares) mice ($n = 4$). Separation score is the $\log_2$ ratio of the exon-skipping probe set intensity versus the inclusion probe set intensity in a mutant relative to wild type (see inset; also see Online Methods). (**b**) Diagram showing numbers of aberrant splicing events in HSA^{LR} mice and $Mbnl1^{\Delta E3/\Delta E3}$ mice with |sepscore| ≥ 0.3. (**c**) Correlation of sepscore values for aberrant splicing in HSA^{LR} compared to $Mbnl1^{\Delta E3/\Delta E3}$ mice (Pearson $R^2 = 0.84$). (**d**–**f**) Validation of members from different classes of events by RT-PCR. Shown are exons either up- (*Nfix*) or downregulated (*Tlk*) in both HSA^{LR} and $Mbnl1^{\Delta E3/\Delta E3}$ mice (**d**) and exons affected only in $Mbnl1^{\Delta E3/\Delta E3}$ (**e**) or only in HSA^{LR} (**f**) mice. The numbers in parentheses represent the length in nucleotides of the affected exon (gray box), shown with flanking exons (white boxes). We quantified the RT-PCR products from three individual mice using the Bioanalyzer and calculated inclusion rates. We judged samples as being different from wild type if a *t*-test indicated that the sample was unlikely to be from the wild-type distribution with $P < 0.05$ (asterisks).

were greatly enriched near the affected exons, suggesting that CUG^{exp} RNA affects splicing primarily through Mbnl1. In contrast to splicing, many changes in mRNA levels are found in CUG^{exp} RNA–expressing muscle but not in the *Mbnl1* mutant, indicating a distinct second defect caused by CTG^{exp} DNA. This defect seems unusually focused on genes expressing extracellular matrix (ECM) components and their regulators, some of which are known to have roles in other forms of muscular dystrophy and connective tissue diseases.

RESULTS

Splicing perturbations in muscle of HSA^{LR} and $Mbnl1^{\Delta E3/\Delta E3}$ mice

If Mbnl1 sequestration is the main cause of gene-expression perturbation in DM1, then transcripts from muscle genetically lacking Mbnl1 should appear broadly similar to those expressing CUG^{exp} RNA. The extent of this similarity should provide a strong quantitative estimate of the extent to which the Mbnl1 sequestration hypothesis can explain the disease. To test this directly, RNA was extracted from the quadriceps muscles of age-matched males either carrying the HSA^{LR} transgene expressing CUG^{exp} RNA (HSA^{LR})[7], homozygous for the *Mbnl1* knockout allele ($Mbnl1^{\Delta E3/\Delta E3}$)[27] or homozygous wild-type *Mbnl1* in the same (FBV) background, and analyzed on splicing-sensitive microarrays as described previously[29,30]. Experience with this method indicates that differences in the $\log_2$ of the ratio of exon skipping to inclusion (the skip/include ratio **Fig. 1**) between two samples (**Fig. 1a**, sepscore; see Online Methods) with absolute value >0.3 can be validated by RT-PCR about 85% of the time[29,30]. We observed 246 events in $Mbnl1^{\Delta E3/\Delta E3}$ mice and 221 events in HSA^{LR} mice that exceeded this score, distributed among different splicing modes (alternative cassette exons, alternative 5′ or 3′ splice sites, and mutually exclusive cassette exons; **Fig. 1b** and **Supplementary Table 1**). Even at this relatively crude level of analysis, nearly 80% of splicing events altered in HSA^{LR} mice were also altered in $Mbnl1^{\Delta E3/\Delta E3}$ mice. About half showed increased exon skipping after loss of Mbnl1, and half showed decreased skipping, indicating formally that Mbnl1 contributes to both activation and repression of splicing.

To evaluate whether the two disease models share quantitatively similar changes in splicing, we compared the sepscores for each splicing event altered (with |sepscore| ≥ 0.3) in both models. The R^2 value for this comparison was 0.84, suggesting that Mbnl1 loss of function explains more than 80% of the splicing phenotype caused by CUG^{exp} RNA expression (**Fig. 1c**). We used RT-PCR to validate a subset of the alternative cassette exons, as shown for a 123-nucleotide cassette exon repressed by Mbnl1 in the *Nfix* gene (encoding NF-I/X CAAT box-binding transcription factor, **Fig. 1d**; see Online Methods). The 93-nucleotide exon of the *Tlk* gene represents the contribution of Mbnl1 to splicing activation, as skipping is increased in the mutants (**Fig. 1d**). Of 33 other events tested, all but 1 behaved as expected from the arrays (FDR = 0.03, **Supplementary Fig. 1** and **Supplementary Table 1**). Of these, 28 are affected in both HSA^{LR} and $Mbnl1^{\Delta E3/\Delta E3}$ mice (**Fig. 1d**; **Supplementary Fig. 1a**); 4 appear affected only in $Mbnl1^{\Delta E3/\Delta E3}$ (such as Ndrg2, **Fig. 1e**; **Supplementary Fig. 1b**); and 1 appears affected only in HSA^{LR} (**Fig. 1f**; **Supplementary Fig. 1c**). The concordance observed by arrays (**Fig. 1c**) was confirmed by RT-PCR reactions quantified on an Agilent Bioanalyzer ($R^2 = 0.88$, **Supplementary Fig. 1d**).

We found four splicing events affected in $Mbnl1^{\Delta E3/\Delta E3}$ but not in HSA^{LR} mice (**Fig. 1e** and **Supplementary Fig. 1b**). These might compete effectively for Mbnl1 against CUG^{exp} RNA and thus may not be inhibited except at levels higher than those achieved by this transgene. We found only one event affected specifically in HSA^{LR} mice (**Fig. 1f**), suggesting that CUG^{exp} RNA transgene expression can cause splicing defects by mechanisms other than simple loss of Mbnl1. Possibly the regulation of this exon is the responsibility of other factors that function in the Mbnl1 knockout mouse but not in HSA^{LR}, such as Mbnl2, which is also sequestered by CUG^{exp} RNA[31,32], or CUGBP1, whose activity is altered in CUG^{exp} RNA–expressing cells[24]; or perhaps the regulation is controlled by some other indirect or combinatorial mechanism (see below). Given that Mbnl2 is sequestered by CUG^{exp} RNA[31,32], our results also indicate that Mbnl2 is largely unable to compensate for the loss of Mbnl1 in muscle: otherwise, we would have found many more splicing events perturbed in the HSA^{LR} but not in the $Mbnl1^{\Delta E3/\Delta E3}$ mouse. We conclude that the vast majority of splicing dysregulation in the CUG^{exp} RNA–expressing muscle is due to catastrophic loss of Mbnl1 splicing-factor activity.

Sequence signatures of Mbnl1-regulated exons in muscle

To identify motifs associated with Mbnl1-regulated exons, we used Improbizer, which identifies sequence motifs present in a set of sequences

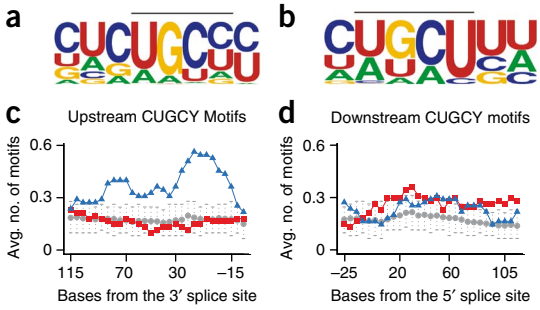

Figure 2 RNA motifs found by bioinformatics analysis near exons altered in the mouse models. (**a**,**b**) Improbizer identifies a motif containing CUGCY upstream of Mbnl1-repressed exons (**a**) and downstream of Mbnl1-activated exons (**b**). (**c**,**d**) Mapping CUGCY elements upstream (**c**) and downstream (**d**) of the Mbnl1-repressed and Mbnl1-activated exons. Each point represents the average frequency of CUGCY element for the 55 Mbnl1-repressed exons (blue triangles), the 66 Mbnl1-activated exons (red squares) or the 790 expressed alternative cassette exons whose gene-expression level did not change in the experiment (gray circles). Error bars indicate ± 2 s.d. of the mean frequency distribution for this population of background exons.

as compared to a background sequence set (see Online Methods). We contrasted the introns upstream and downstream of three groups of exons (**Supplementary Table 2**): Mbnl1-activated exons (more skipping in *Mbnl1*$^{\Delta E3/\Delta E3}$ than in wild type); Mbnl1-repressed exons (more inclusion in *Mbnl1*$^{\Delta E3/\Delta E3}$); and background exons (detected in our experiments but not showing substantial splicing change). Improbizer strongly recognizes enrichment of a YGCY-containing motif (CUGCY) in the Mbnl1-repressed and Mbnl1-activated exons (**Fig. 2a**,**b**), consistent with interpretation of mutagenesis experiments on several splicing substrates[18,33,34] and with crystal complexes of Mbnl1 zinc fingers with RNA[20]. We mapped the distribution of positions of CUGCY motifs to upstream and downstream intron regions (**Fig. 2c**,**d**). Introns upstream of Mbnl1-repressed exons, as well as part of the exon itself, were strongly enriched for CUGCY motifs, which often occur multiple times in individual pre-mRNAs. In contrast, the corresponding region of Mbnl1-activated exons was not enriched (**Fig. 2c**). In the intron downstream of affected exons, both Mbnl1-repressed and Mbnl1-activated exons showed slight but marked enrichment in CUGCY elements above the background exons (**Fig. 2d**). Enrichment for CUGBP1 or hnRNP H motifs was not observed (**Supplementary Fig. 2**). This pattern supports the argument that the splicing defects are primarily due to loss of Mbnl1 and suggests that regulation by Mbnl1 often involves its direct binding to regions near the exons it regulates.

To confirm the function of YGCY motifs in newly identified exons, we cloned the Mbnl1-activated exon from the *Vldlr* gene and the Mbnl1-repressed exon from the *Nfix* gene, each with their native flanking intron sequences, into a splicing reporter plasmid. We then mutated copies of the motif in each and tested them in mouse embryo fibroblasts lacking Mbnl1 (derived from the *Mbnl1*$^{\Delta E3/\Delta E3}$ mouse), along with a construct expressing an Mbnl1-GFP fusion protein or GFP alone (**Fig. 3a**,**b**). Splicing activation is promoted by the Mbnl1-GFP fusion protein for the wild-type but not the mutant *Vldlr* splicing reporter, indicating that the motif downstream of the exon is important for Mbnl1-mediated splicing activation (**Fig. 3a**). Splicing repression is promoted by the Mbnl1-GFP fusion protein for the wild-type *Nfix* reporter and is only partly compromised by the alteration of four copies of the motif in the intron upstream of the exon. Complete loss of repression is achieved once three additional motif copies in the exon itself are altered (**Fig. 3b**). These results confirm by reconstruction that newly identified exons depend on Mbnl1 for their correct splicing, and furthermore show that regulation is mediated by sequence motifs enriched in Mbnl1-regulated exons.

Mis-splicing events found in mouse are observed in human DM1

A handful of splicing defects are known in humans with DM1 (ref. 11). To determine whether humans with DM1 share newly identified splicing changes with the mouse DM1 models, we aligned affected mouse exons to the human genome and found 49 that are conserved and orthologous to Mbnl1-dependent mouse exons, 39 of which show strong evidence of alternative splicing in humans, according to the Alt Events track on the University of California, Santa Cruz Genome Browser[35]. We tested six for perturbation of splicing in people with DM1. The splicing of three genes (*NFIX*, *SMYD1* and *SPAG9*) was perturbed in all three DM1-affected individuals tested, as compared to an unaffected human control, whereas for three others (*GNAS*, *MTDH* and *PPP2R5C*) splicing was clearly affected only in one or two of the three affected individuals (**Fig. 4**). To rigorously determine the value of such changes in monitoring human DM1, more DM1-affected and unaffected individuals would need to be tested. Nonetheless, these and other[11] splicing changes indicate that Mbnl1-dependent events observed in the mouse are also observed in human DM1. Thus, exons affected in the mouse models

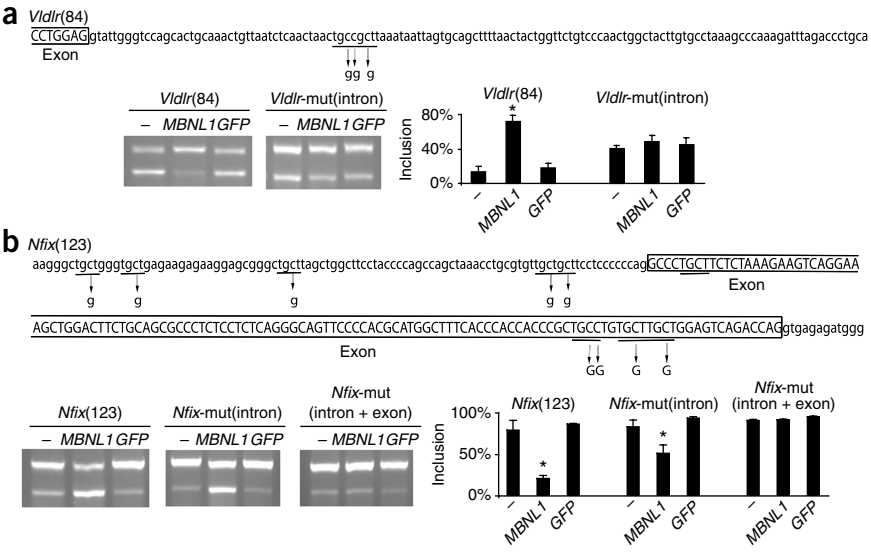

Figure 3 YGCY motifs mediate Mbnl1-dependent splicing repression and activation. (**a**,**b**) Intron (lower case) and exon (upper case and boxed) sequences for the *Vldlr*(84) (**a**) and *Nfix*(123) (**b**) exons are shown, and mutations that disrupt YGCY motifs are labeled with arrows. We transfected mouse embryo fibroblasts lacking endogenous Mbnl1 (derived from the *Mbnl1*$^{\Delta E3/\Delta E3}$ mouse) with the splicing reporter and either no expression plasmid (–), a plasmid producing Mbnl1-GFP protein (*MBNL1*) or just GFP protein (*GFP*), and we measured exon inclusion by RT-PCR using the Bioanalyzer. Significant splicing changes (*t*-test–derived $P < 0.05$) are labeled with an asterisk.

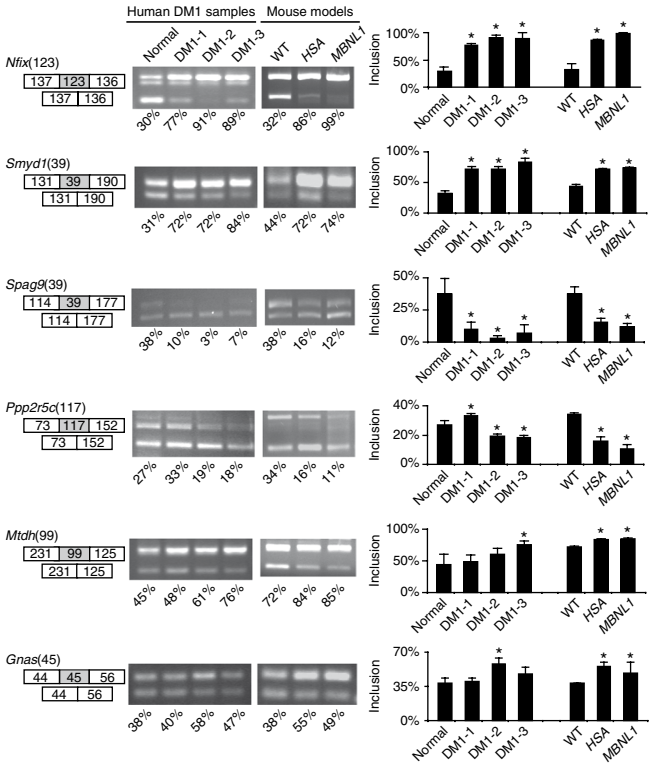

Figure 4 Test of humans with DM1 for splicing perturbations newly predicted from mouse model microarray data. We compared skeletal muscle RNA from three individuals with DM1 and one unaffected human control to those from HSA^{LR} and $Mbnl1^{\Delta E3/\Delta E3}$ mice by RT-PCR. We determined inclusion rates on a Bioanalyzer. Statistically significant differences from the unaffected individual human are indicated by an asterisk; however, this captures only technical variation in the RT-PCR protocol, as a single original RNA donation was processed in triplicate. For mouse samples, we compared three individuals of each genetic type, so statistical differences here arise from both biological and technical variation. Note that extent of genetic variation between the humans is unknown (except at the CTG^{exp} locus) but is likely to be much greater than the differences between the mouse groups.

represent a rich source of potential markers for detecting and evaluating the human disease as well as for testing the efficacy of treatments.

Many changes in HSA^{LR} mice are not explained by loss of Mbnl1

Besides influencing splicing, loss of Mbnl1 could lead to transcript-level changes, either directly, through nonsense-mediated decay caused by aberrant splicing (AS-NMD)[36] or altered mRNA stability, or indirectly, through mis-splicing of transcription-factor or splicing-factor mRNAs, all of which could combine to alter the transcriptome[37]. If Mbnl1 sequestration is the only mechanism by which CTG^{exp} DNA alters gene expression, these myriad effects should be highly similar in both mouse models. To test this, we compared gene-expression changes in HSA^{LR} mice to those in $Mbnl1^{\Delta E3/\Delta E3}$ mice using the array probe intensities for all the constitutive (always-included) exons in each gene (see Online Methods). After filtering out genes with probes affected by cross-hybridization to the massive amounts of CUG^{exp} RNA in the HSA^{LR} mouse samples (see Online Methods), we identified 148 genes whose muscle transcript levels were significantly altered in HSA^{LR} mice and 110 in the $Mbnl1^{\Delta E3/\Delta E3}$ mice (≥1.5-fold change; $q < 0.05$ cutoff, **Fig. 5a**). Of the 148 changes observed in HSA^{LR} mice, 102 (69%) also appeared in $Mbnl1^{\Delta E3/\Delta E3}$ mice, suggesting that, as for splicing, the underlying reason for those expression changes is loss of Mbnl1, either directly or indirectly through other factors. The 46 genes (31%) that change only in HSA^{LR} mice indicate that CTG^{exp} DNA imposes a second layer of dysregulation not shared by mice lacking Mbnl1. Overall, Mbnl1 loss explains a little more than half of the transcript-level variance between wild-type and HSA^{LR} mice (**Fig. 5b**; $R^2 = 0.57$). We named the larger class of genes whose mRNA levels are altered in both models primarily because of loss of Mbnl1 "class I" genes, and those altered only in the HSA^{LR} but not $Mbnl1^{\Delta E3/\Delta E3}$ mice "class II" genes. In a more complex study using a very different array platform[37], similar gene classes were identified that contained different genes (see Discussion).

To validate the transcript-level changes, we used quantitative RT-PCR (qPCR, **Fig. 5c** and **Supplementary Table 3**). Among the changes observed only in the HSA^{LR} mice, more genes are downregulated than upregulated (**Fig. 5c**). There is good agreement with the microarray data; however, the magnitude of the change detected is larger by qPCR, likely because of the smaller dynamic range of arrays. Because the class II changes do not occur in the $Mbnl1^{\Delta E3/\Delta E3}$ mouse (at least by 12–14 weeks of age), they are unlikely to be a simple consequence of a loss of Mbnl1. However, the HSA^{LR} mouse also suffers a loss of Mbnl1 function[28], and it remains possible that Mbnl1 loss is necessary but not sufficient to produce the class II changes. We conclude that a substantial number of transcript-level changes occur in the HSA^{LR} mouse that do not occur in the $Mbnl1^{\Delta E3/\Delta E3}$ mouse.

To capture class II gene functions, we relaxed the array-measured fold change cutoff from 1.5- to 1.3-fold, since qPCR indicated that fold change is underestimated in this data. Because the filtering of cross-hybridizing CUG^{exp} RNA in the samples is incomplete, some genes will artificially appear to be upregulated in the HSA^{LR} mouse but not in the $Mbnl1^{\Delta E3/\Delta E3}$ mouse. We avoided such genes by only considering those that were significantly downregulated in the HSA^{LR} mouse relative to wild type ($q < 0.05$). To ensure that the genes on the list were specific to CTG^{exp} effects, we required that their q values (FDRs) in both the HSA^{LR} versus wild-type and the HSA^{LR} versus

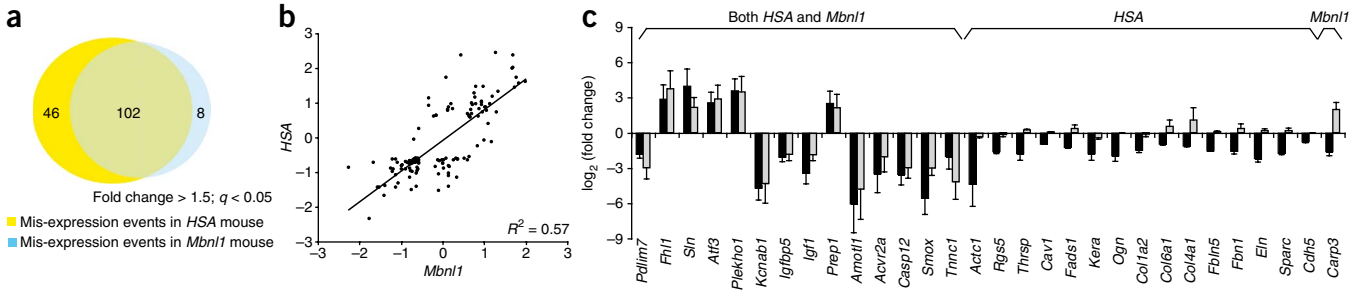

Figure 5 Comparison of altered mRNA levels in the two mouse models. (**a**) Numbers of genes with altered mRNA levels in HSA^{LR} and $Mbnl1^{\Delta E3/\Delta E3}$ mice (cutoff: fold change ≥ 1.5; $q \leq 0.05$). (**b**) Comparison of magnitudes of the $\log_2$(fold change) in mRNA levels in the two mouse models (Spearman $R^2 = 0.57$). (**c**) Validation of altered mRNA levels found on arrays by quantitative RT-PCR. Quantification of triplicate data with error bars (± s.d.) is shown, where black bars represent change in HSA^{LR} mice, and gray bars represent change in $Mbnl1^{\Delta E3/\Delta E3}$ mice.

Table 1 Extracellular matrix and disease associations of class II genes

Gene	Protein name	Function in ECM and cell adhesion	Human disease or mouse knockout phenotype
COL15A1	Procollagen, type XV	Adhesion function in skeletal and cardiac muscle	Col15a1-deficient mice have skeletal myopathy and cardiovascular defects
COL1A1	Procollagen, type I, α 1	Abundant, ECM organization	Osteogenesis imperfecta Type I; Ehlers-Danlos syndrome (EDS)
COL1A2	Procollagen, type I, α 2	Abundant, ECM organization	Osteogenesis imperfecta Type I
COL4A1	Procollagen, type IV, α 1	ECM organization	Familial porencephaly; Alport syndrome
COL6A1	Procollagen, type VI, α 1	Anchors the muscle basement to the ECM	Ullrich congenital muscular dystrophy; Bethlem myopathy
ELN	Elastin	ECM organization	Supravalvular aortic stenosis; Williams-Beuren syndrome
ENG	Endoglin (TGF-β type III receptor)	Interacts with TβR-II. Cytoplasmic domain phosphorylated by ALK kinases.	Hereditary hemorrhagic telangiectasia of Rendu, Osler and Weber
FBLN5	Fibulin 5	Elastin fiber formation	Autosomal recessive cutis laxa, type I; age-related muscular degeneration
FBN1	Fibrilin-1	Elastin fiber assembly and formation	Marfan syndrome
FMOD	Fibromodulin	Inhibits type I collagen fibrillogenesis	Ehlers-Danlos syndrome, type I
KERA	Keratocan	Keratan sulfate proteoglycan of the extracellular matrix; keratocan expression in mice is limited to the cornea in the adult organism	Cornea plana 2
LOX	Lysyl oxidase	Formation and repair of the ECM by oxidizing lysine residues in elastin and collagen	Cardiovascular dysfunction in *Lox* knockout mice
PLOD1	Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 1	Lysyl hydroxylase that controls cross-links	Ehlers-Danlos syndrome type VIA; Nevo syndrome
POSTN	Periostin, osteoblast specific factor	Regulation of collagen fiber diameter and cross-linking	Cardiac valve disease in *Postn* knockout mice have
TNC	Tenascin C	Regulation of muscle strength	Weak muscles in *Tnc*-deficient mice

$Mbnl1^{\Delta E3/\Delta E3}$ mouse comparisons be <0.05 and that their q values in the Mbnl1 versus wild-type comparisons be >0.05. This resulted in a set of 93 genes downregulated in the HSA^{LR} mouse but not changed in the $Mbnl1^{\Delta E3/\Delta E3}$ mouse (**Supplementary Table 4**).

Distinct impacts of CUG^{exp} imposed through different sets of genes

Using the splicing changes and the class I and class II gene-expression change lists, we searched for functionally related gene classes that might provide insight into the disease. We performed a Gene Ontology (GO) analysis[38] using the background set of all genes expressed in the experiment (~6,000 genes of the ~10,000 on the array). Splicing changes were not significantly enriched ($q \geq 0.05$) in any GO category (data not shown), suggesting that Mbnl1-regulated exons are distributed in several broad functional classes of genes. The category of class I genes, for which changes in expression presumably arise as an indirect consequence of the loss of Mbnl1 in splicing, is enriched for genes encoding insulin and insulin-like growth factor signaling and glucose metabolism (**Supplementary Table 5**), suggesting that loss of Mbnl1 has strong but specific effect on glucose metabolism in muscle cells. Other muscle-cell terms such as contractile fiber and muscle development are also strongly associated with this gene set.

Class II genes are very significantly enriched for those encoding or regulating components of the ECM ($q < 0.001$), with one-fifth of the genes (18/93) carrying this functional label (**Supplementary Table 6**). Fifteen are orthologous to human genes involved in other forms of muscular dystrophy, myopathies and several connective tissue diseases, or their deletion in mouse creates a similar defect (**Table 1**, **Supplementary Table 7**). We conclude that the coordinated dysregulation of a large set of genes associated with ECM function is caused by a second effect of CTG^{exp} DNA that is not explained by the simple loss of Mbnl1.

We analyzed the sequences associated with the 5′ and 3′ UTRs of the downregulated class II genes. Using the frequency of 7-mer sequence words, we find that sequences containing the Mbnl1 motif (YGCY) are enriched in the UTRs of the 93 class II mRNAs. There is a substantial enrichment of GGUGCUA in the 3′ UTRs as well as enrichment of several related words in the 5′ UTRs (UGCCUGC, CCUGCCU, UGUGCCU; **Supplementary Table 8**). Because of the extreme similarity between the RNA-binding domains of Mbnl1 and Mbnl2, it seems likely that the two proteins bind highly related RNA sequences[21,39]. Because Mbnl2 is sequestered by CUG^{exp} RNA[31,32], these results lead to the hypothesis that expression of class II genes might be altered through loss of Mbnl2 function mediated through binding to the UTRs of the mRNAs. The enrichment of Mbnl binding sites observed in the 93 class II mRNAs is also observed in the subgroup of class II genes that encode ECM components (**Supplementary Table 8**). Mbnl2 colocalizes with integrin α3 mRNA in the cytoplasm and promotes its expression at the cell surface[40]. Because integrins are ECM components, it is tempting to propose that Mbnl2 plays a general role in promoting correct localization, stability and translation of mRNAs encoding ECM proteins and their regulators.

DISCUSSION

Loss of Mbnl1 is the major cause of CUG^{exp} RNA splicing toxicity

Our genome-wide test of the Mbnl1 sequestration hypothesis identified a large set of mouse muscle splicing events that depend on Mbnl1 and compared them to splicing changes caused by toxic CUG^{exp} RNA. Strikingly, genetic loss of Mbnl1 accounts for about 80–90% of the altered splicing phenotype in the CUG^{exp} RNA–expressing mice (**Fig. 1** and **Supplementary Fig. 1**). The extreme similarity of global splicing defects supports a sequestration model in which CUG^{exp} RNA creates a catastrophic loss of Mbnl1 splicing function in DM1. This is confirmed by the presence of Mbnl1 binding motifs in and near exons whose splicing is compromised by loss of Mbnl1 (**Figs. 2** and **3**). Our analysis was unable to identify enrichment of motifs for other splicing factors implicated in DM1 pathogenesis, such as CUGBP1

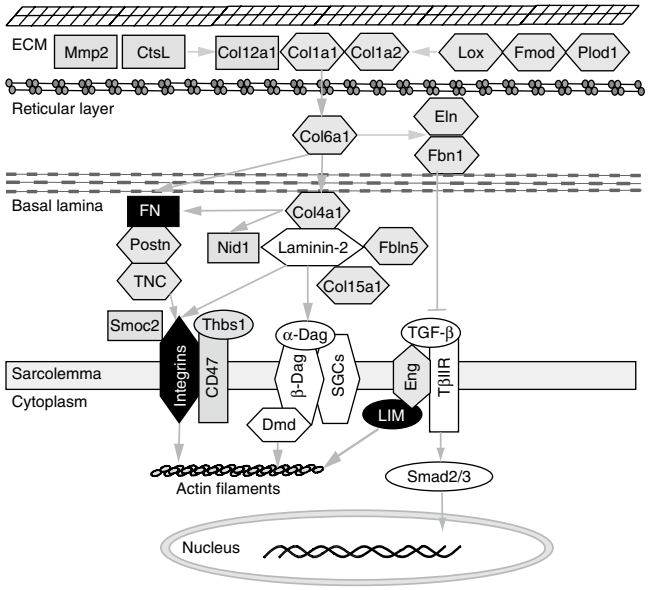

Figure 6 ECM proteins whose mRNA levels are altered by CUG^{exp} RNA expression. The proteins whose mRNAs are misregulated in *HSA*^{LR} but not *Mbnl1*$^{\Delta E3/\Delta E3}$ mice (class II genes) are shown in gray, and other proteins in the same network are shown in white. Proteins in black are those whose mRNAs have splicing perturbations in both *HSA*^{LR} and *Mbnl1*$^{\Delta E3/\Delta E3}$ mice. Proteins in hexagons are known to be associated with other muscular dystrophies or connective tissue diseases (see text), and gray arrows denote regulatory interactions.

(refs. 23,25) or hnRNP H (ref. 23), and neither we (this study) nor previous work[25] studying the developing heart found strong coenrichment of CUGBP1 motifs with Mbnl1-dependent exons. We also found rare splicing changes specific to each model (**Fig. 1**). Thus, although Mbnl1 loss is the major contributor to splicing defects in the *HSA*^{LR} mouse, other factors also contribute in more subtle ways.

A cascade of other perturbations due to Mbnl1 loss also occurs. These encompass changes in transcript level (**Fig. 5**; see also ref. 37). Fully 69% of the transcript-level changes observed in the *Mbnl1*$^{\Delta E3/\Delta E3}$ mouse are also observed in the *HSA*^{LR} mouse (class I genes). Many of these involve no altered splicing, but they could depend directly on Mbnl1 (if it has a role in mRNA stabilization) or could be altered through indirect effects resulting from the loss of splicing regulation for transcription or other RNA stability factors. The broad impact of Mbnl1 loss makes following the threads of each perturbation to the mechanistic source of each disease symptom a daunting prospect. Whether direct or indirect, perturbations that arise from loss of Mbnl1 are pervasive and complex and may explain the medical complexities of the human disease. The mouse models reveal defects parallel to those in humans with DM1 (**Fig. 4**), suggesting that diagnostic tests for a wide spectrum of genes that depend on Mbnl1 function may presage DM1 onset or allow sensitive evaluation of therapies.

What causes perturbations not attributable to Mbnl1 loss?

A second class of gene-expression perturbations occurs in CUG^{exp} RNA–expressing mice but not in mice lacking Mbnl1 (class II genes). A surprising fraction is associated with ECM function (**Fig. 6** and **Supplementary Tables 6** and **7**). We propose that their dysregulation is due to loss of function of a second factor. The leading candidate for such a factor is Mbnl2, which like Mbnl1 is sequestered by CUG^{exp} RNA[5,31,32,41]. The RNA-binding domains of Mbnl1 and Mbnl2 are nearly identical, and UTRs of class II genes are enriched in Mbnl RNA motifs (**Supplementary Table 8**), supporting this hypothesis. A mouse Mbnl2 mutant has a relatively mild phenotype that includes muscle fibrosis; however, expression of residual Mbnl2 has not been strictly ruled out in this system[42]. The enrichment of ECM genes among the class II set (**Fig. 6** and **Supplementary Tables 6** and **7**), combined with the finding that Mbnl2 colocalizes with integrin α3 mRNA and promotes localization of the integrin protein at focal adhesions[40], makes it reasonable to postulate that Mbnl2 serves a general role in promoting correct expression of ECM molecules.

Tracing CUG^{exp} RNA toxicity from gene to disease phenotype

Loss of Mbnl1 through sequestration by CUG^{exp} RNA contributes to changes in muscle morphology and excitability as indicated by studies with Mbnl1-deficient mice[27]. These include aberrant splicing of *Clcn1*, causing myotonia[14,15,43]; mis-splicing of *Camk2g*, causing interference in excitation-contraction coupling[44]; and alteration of *Acvr2a* expression, disturbing regulation of muscle growth[45]. A large number of new candidates associated with ion channels, cytoskeleton and myotubule composition, glucose metabolism and signal transduction may lead to additional defects (**Supplementary Tables 1** and **3**).

Despite this, loss of Mbnl1 does not produce the progressive muscle wasting observed in DM1. Previous work reported that transgenic mice with 960 CTG repeats in the *Dmpk* 3′ UTR display muscle loss not observed in *Mbnl1*$^{\Delta E3/\Delta E3}$ mice[6,27]. We found that CUG^{exp} RNA but not genetic loss of Mbnl1 causes alteration of a broad set of ECM mRNAs (**Fig. 6**) that could worsen with age or increased repeat expansion, as observed in the human disease. Consistent with this, mutations in some of these same ECM genes cause muscular dystrophies or connective tissue diseases[46–48] (**Table 1**). For example, defects in collagen I cause osteogenesis imperfecta with muscle weakness[49], whereas mutations in Col6a1 are responsible for Ullrich congenital muscular dystrophy[50]. Col15a1 is involved in maintaining ECM structural integrity, and Col15a1-deficient mice show progressive muscle fiber degeneration[51]. Defects in elastin assembly also affect muscle strength and regeneration. Fibrillin-1 (*FBN1*) mutation causes Marfan syndrome with muscle involvement[52,53]. Loss of fibulin-5 contributes to age-related muscular degeneration[54,55]. Tenascin C interacts with integrin (integrin α7 mutations cause congenital muscular dystrophy), and *Tnc*-deficient mice show reduced muscle strength[56]. Unlike EpA960 transgenic mice[6], *HSA*^{LR} mice do not develop severe muscle weakness and wasting, so more extreme alterations in ECM gene expression may promote a more severe dystrophic phenotype. Together, these data overwhelmingly point to the idea that a loss of regulation of ECM function and the consequent effects on cell adhesion contribute to muscle defects in DM1 (**Table 1** and **Fig. 6**).

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

Accession codes. Microarray data for this study has been deposited in GEO under accession number GSE17986.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank D. Black (Univ. of California, Los Angeles) and X.-D. Fu (Univ. of California, San Diego) for gifts of plasmids and advice on mammalian cell culture. Thanks to B. Chabot and J. Sanford for critical reading of the manuscript. This work was primarily supported by US National Institutes of Health grant GM084317 to M.A., with additional support from US National Institutes of Health grant

GM040478 (to M.A.), US National Institutes of Health grant AR046799 (to M.S.S.) and US National Institutes of Health grant AR046806 and Muscular Dystrophy Center grant U54NS048843 (to C.A.T.). M.P.H. acknowledges the California Institute of Regenerative Medicine for postdoctoral support.

AUTHOR CONTRIBUTIONS

M.A., H.D., M.S.S. and C.A.T. designed the experiments; H.D., R.J.O., D.L.T. and M.P.H. performed the experiments; T.A.C. provided materials and data analysis; M.S.C., J.P.D. and L.S. performed data analysis; M.A., H.D. and M.S.C. wrote the paper.

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

Published online at http://www.nature.com/nsmb/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Mirkin, S.M. Expandable DNA repeats and human disease. *Nature* **447**, 932–940 (2007).
2. Caskey, C.T., Pizzuti, A., Fu, Y.H., Fenwick, R.G. Jr. & Nelson, D.L. Triplet repeat mutations in human disease. *Science* **256**, 784–789 (1992).
3. Buxton, J. *et al.* Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. *Nature* **355**, 547–548 (1992).
4. Taneja, K.L., McCurrach, M., Schalling, M., Housman, D. & Singer, R.H. Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J. Cell. Biol.* **128**, 995–1002 (1995).
5. 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).
6. Orengo, J.P. *et al.* Expanded CTG repeats within the DMPK 3′ UTR causes severe skeletal muscle wasting in an inducible mouse model for myotonic dystrophy. *Proc. Natl. Acad. Sci. USA* **105**, 2646–2651 (2008).
7. Mankodi, A. *et al.* Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science* **289**, 1769–1773 (2000).
8. Ebralidze, A., Wang, Y., Petkova, V., Ebralidse, K. & Junghans, R.P. RNA leaching of transcription factors disrupts transcription in myotonic dystrophy. *Science* **303**, 383–387 (2004).
9. Krol, J. *et al.* Ribonuclease dicer cleaves triplet repeat hairpins into shorter repeats that silence specific targets. *Mol. Cell* **25**, 575–586 (2007).
10. Kuyumcu-Martinez, N.M., Wang, G.S. & Cooper, T.A. Increased steady-state levels of CUGBP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation. *Mol. Cell* **28**, 68–78 (2007).
11. Lin, X. *et al.* Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy. *Hum. Mol. Genet.* **15**, 2087–2097 (2006).
12. Lueck, J.D., Mankodi, A., Swanson, M.S., Thornton, C.A. & Dirksen, R.T. Muscle chloride channel dysfunction in two mouse models of myotonic dystrophy. *J. Gen. Physiol.* **129**, 79–94 (2007).
13. Charlet, B.N. *et al.* Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. *Mol. Cell* **10**, 45–53 (2002).
14. Lueck, J.D. *et al.* Chloride channelopathy in myotonic dystrophy resulting from loss of posttranscriptional regulation for CLCN1. *Am. J. Physiol. Cell Physiol.* **292**, C1291–C1297 (2007).
15. Wheeler, T.M., Lueck, J.D., Swanson, M.S., Dirksen, R.T. & Thornton, C.A. Correction of ClC-1 splicing eliminates chloride channelopathy and myotonia in mouse models of myotonic dystrophy. *J. Clin. Invest.* **117**, 3952–3957 (2007).
16. Savkur, R.S., Philips, A.V. & Cooper, T.A. Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. *Nat. Genet.* **29**, 40–47 (2001).
17. Miller, J.W. *et al.* Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *EMBO J.* **19**, 4439–4448 (2000).
18. Warf, M.B. & Berglund, J.A. MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. *RNA* **13**, 2238–2251 (2007).
19. Begemann, G. *et al.* muscleblind, a gene required for photoreceptor differentiation in *Drosophila*, encodes novel nuclear Cys3His-type zinc-finger-containing proteins. *Development* **124**, 4321–4331 (1997).
20. Teplova, M. & Patel, D.J. Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1. *Nat. Struct. Mol. Biol.* **15**, 1343–1351 (2008).
21. Ho, T.H. *et al.* Muscleblind proteins regulate alternative splicing. *EMBO J.* **23**, 3103–3112 (2004).
22. Mooers, B.H., Logue, J.S. & Berglund, J.A. The structural basis of myotonic dystrophy from the crystal structure of CUG repeats. *Proc. Natl. Acad. Sci. USA* **102**, 16626–16631 (2005).
23. Paul, S. *et al.* Interaction of muscleblind, CUG-BP1 and hnRNP H proteins in DM1-associated aberrant IR splicing. *EMBO J.* **25**, 4271–4283 (2006).
24. Timchenko, N.A. *et al.* RNA CUG repeats sequester CUGBP1 and alter protein levels and activity of CUGBP1. *J. Biol. Chem.* **276**, 7820–7826 (2001).
25. Kalsotra, A. *et al.* A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. *Proc. Natl. Acad. Sci. USA* **105**, 20333–20338 (2008).
26. Kim, D.H. *et al.* HnRNP H inhibits nuclear export of mRNA containing expanded CUG repeats and a distal branch point sequence. *Nucleic Acids Res.* **33**, 3866–3874 (2005).
27. Kanadia, R.N. *et al.* A muscleblind knockout model for myotonic dystrophy. *Science* **302**, 1978–1980 (2003).
28. Kanadia, R.N. *et al.* Reversal of RNA missplicing and myotonia after muscleblind overexpression in a mouse poly(CUG) model for myotonic dystrophy. *Proc. Natl. Acad. Sci. USA* **103**, 11748–11753 (2006).
29. Sugnet, C.W. *et al.* Unusual intron conservation near tissue-regulated exons found by splicing microarrays. *PLOS Comput. Biol.* **2**, e4 (2006).
30. Ni, J.Z. *et al.* Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.* **21**, 708–718 (2007).
31. Fardaei, M. *et al.* Three proteins, MBNL, MBLL and MBXL, co-localize *in vivo* with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum. Mol. Genet.* **11**, 805–814 (2002).
32. Holt, I. *et al.* Muscleblind-like proteins: similarities and differences in normal and myotonic dystrophy muscle. *Am. J. Pathol.* **174**, 216–227 (2009).
33. Hino, S. *et al.* Molecular mechanisms responsible for aberrant splicing of SERCA1 in myotonic dystrophy type 1. *Hum. Mol. Genet.* **16**, 2834–2843 (2007).
34. Yuan, Y. *et al.* Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. *Nucleic Acids Res.* **35**, 5474–5486 (2007).
35. Hsu, F. *et al.* The UCSC known genes. *Bioinformatics* **22**, 1036–1046 (2006).
36. Lejeune, F. & Maquat, L.E. Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr. Opin. Cell Biol.* **17**, 309–315 (2005).
37. Wheeler, T.M. *et al.* Reversal of RNA dominance by displacement of protein sequestered on triplet repeat RNA. *Science* **325**, 336–339 (2009).
38. Zeeberg, B.R. *et al.* High-Throughput GoMiner, an 'industrial-strength' integrative gene ontology tool for interpretation of multiple-microarray experiments, with application to studies of Common Variable Immune Deficiency (CVID). *BMC Bioinformatics* **6**, 168 (2005).
39. He, F. *et al.* Solution structure of the RNA binding domain in the human muscleblind-like protein 2. *Protein Sci.* **18**, 80–91 (2009).
40. Adereth, Y., Dammai, V., Kose, N., Li, R. & Hsu, T. RNA-dependent integrin α3 protein localization regulated by the Muscleblind-like protein MLP1. *Nat. Cell Biol.* **7**, 1240–1247 (2005).
41. Jiang, H., Mankodi, A., Swanson, M.S., Moxley, R.T. & Thornton, C.A. Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Hum. Mol. Genet.* **13**, 3079–3088 (2004).
42. Hao, M. *et al.* Muscleblind-like 2 (Mbnl2)-deficient mice as a model for myotonic dystrophy. *Dev. Dyn.* **237**, 403–410 (2008).
43. Mankodi, A. *et al.* 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 (2002).
44. Xu, X. *et al.* ASF/SF2-regulated CaMKIIdelta alternative splicing temporally reprograms excitation-contraction coupling in cardiac muscle. *Cell* **120**, 59–72 (2005).
45. Lee, S.J. *et al.* Regulation of muscle growth by multiple ligands signaling through activin type II receptors. *Proc. Natl. Acad. Sci. USA* **102**, 18117–18122 (2005).
46. Kanagawa, M. & Toda, T. The genetic and molecular basis of muscular dystrophy: roles of cell-matrix linkage in the pathogenesis. *J. Hum. Genet.* **51**, 915–926 (2006).
47. Jimenez-Mallebrera, C., Brown, S.C., Sewry, C.A. & Muntoni, F. Congenital muscular dystrophy: molecular and cellular aspects. *Cell. Mol. Life Sci.* **62**, 809–823 (2005).
48. Schessl, J., Zou, Y. & Bonnemann, C.G. Congenital muscular dystrophies and the extracellular matrix. *Semin. Pediatr. Neurol.* **13**, 80–89 (2006).
49. Engelbert, R.H. *et al.* Osteogenesis imperfecta in childhood: impairment and disability. A prospective study with 4-year follow-up. *Arch. Phys. Med. Rehabil.* **85**, 772–778 (2004).
50. Ishikawa, H. *et al.* Ullrich disease due to deficiency of collagen VI in the sarcolemma. *Neurology* **62**, 620–623 (2004).
51. Eklund, L. *et al.* Lack of type XV collagen causes a skeletal myopathy and cardiovascular defects in mice. *Proc. Natl. Acad. Sci. USA* **98**, 1194–1199 (2001).
52. Behan, W.M. *et al.* Muscle fibrillin deficiency in Marfan's syndrome myopathy. *J. Neurol. Neurosurg. Psychiatry* **74**, 633–638 (2003).
53. Percheron, G. *et al.* Muscle strength and body composition in adult women with Marfan syndrome. *Rheumatology (Oxford)* **46**, 957–962 (2007).
54. Stone, E.M. *et al.* Missense variations in the fibulin 5 gene and age-related macular degeneration. *N. Engl. J. Med.* **351**, 346–353 (2004).
55. Lotery, A.J. *et al.* Reduced secretion of fibulin 5 in age-related macular degeneration and cutis laxa. *Hum. Mutat.* **27**, 568–574 (2006).
56. Morellini, F. & Schachner, M. Enhanced novelty-induced activity, reduced anxiety, delayed resynchronization to daylight reversal and weaker muscle strength in tenascin-C-deficient mice. *Eur. J. Neurosci.* **23**, 1255–1268 (2006).

ONLINE METHODS

Sample preparation and Affymetrix splicing microarray detection. We compared RNA samples from the quadriceps muscle of individual 12- to 14-week-old male mice from the *Mbnl1*$^{\Delta E3/\Delta E3}$, *HSA*^{LR} and FVB inbred background lines ($n = 4$ for each group). To identify Mbnl1-dependent splicing events in heart (shown in **Supplementary Table 2**), we compared RNA from quadriceps muscle and heart from *Mbnl1*$^{\Delta E3/\Delta E3}$ mice and age-matched wild-type mice in the C57BL/6J background. We processed RNA samples for hybridization to Affymetrix "A-chip" oligonucleotide microarrays[29] according to the standards of the manufacturer.

Microarray data analysis. We performed the analysis as described[29] with modifications indicated below. Preliminary analysis of data for the *HSA*^{LR} mice indicated that probes with any 7-mer comprised of CTG repeats reported much higher intensities than others in the same gene, likely reflecting cross-hybridization driven by the large mass of CUG^{exp} RNA in the sample. We corrected for this by ignoring any gene with more than four probes containing the three 7-mer permutations of $(CTG)_n$: CTGCTGC, GCTGCTG or TGCTGCT.

The separation score (sepscore) method[29] is illustrated for the *Nfix*(123) exon in **Figure 1a**: when exon skipping (Skip) intensities are plotted against exon inclusion (Include) intensities, the wild-type samples appear on the left side of the graph whereas the *Mbnl1*$^{\Delta E3/\Delta E3}$ and *HSA*^{LR} samples appear in the lower right, indicating a shift from exon skipping in the wild-type to inclusion in the mutants. The equation for sepscore is:

$$\text{sepscore} = \log_2[\text{Mut(Skip/Include)/WT(Skip/Include)}] \quad (1)$$

For each replicate set (*HSA*^{LR}, *Mbnl1*$^{\Delta E3/\Delta E3}$ and wild type), we estimated the $\log_2$ ratio of skipping to inclusion using robust least-squares analysis. We evaluated sepscore significance by permuting the assignments of data points to replicate sets, calculating the sepscore for the permuted data and estimating the likelihood that the observed data came from the permuted distribution. We estimated rates of overall gene expression according to probe sets that measure constitutive features. We identified genes that are differentially expressed between replicate sets using the SAM software[57].

RT-PCR and Bioanalyzer analysis. We reverse transcribed total RNA to cDNA with Superscript III (Invitrogen) and oligo-dT using the manufacturer's protocol. Then we used ~50 ng cDNA as template for PCR with primers to regions spanning the test cassette exons (**Supplementary Table 9**). We carried out PCR for 25–35 cycles using the Platinum Taq polymerase (Invitrogen). To measure splicing, we separated 1 μl PCR products from above reaction on the Agilent Bioanalyzer, which then reported the size and concentrations of each splicing-derived product.

Quantitative RT-PCR. We carried out quantitative RT-PCR in 30-μl reactions using Brilliant QRT-PCR Master Mix kits (SYBR-Green, Stratagene) on a Bio-Rad iCycler. We used GAPDH for normalization and the protocol of previous work[58] for calculating changes in threshold cycles.

Analysis of sequence motifs. We used Improbizer to search intronic sequences flanking each differentially-spliced exon. Background sequences were from alternative exons that were expressed but showed no changes in splicing. We searched for four copies of each motif during its search (maxOcc setting = 4). To map motifs, we searched in 40-base windows (offset by 5 bases) through the 150-base region immediately upstream and downstream of each alternative exon. We contrasted 55 Mbnl1-repressed exons (skipping lost in the mutant), 66 Mbnl1-induced exons (inclusion lost in the mutant) and 790 alternative exons expressed without splicing changes. We determined background motif frequencies by randomly sampling equivalent numbers of exons from the background 1,000 times per sequence window. We determined mean background counts and the 95% confidence interval through quartile analysis.

To identify sequence words associated with class II genes, we contrasted the sequences of the 5′ and 3′ UTRs to those of all genes expressed in the experiment. We determined the number of times each k-mer appeared in the selected sequences and in the background, for $k = 5$, $k = 6$ and $k = 7$. We used a public Perl module (Statistics::ChisqIndep) for χ^2 evaluation of 2×2 contingency table data to determine the likelihood (p value) that the observed enrichment was due to chance. We applied a Bonferroni correction by multiplying the p value by the number of all possible k-mers (4^k). This correction does not assume independence and ignores the fact that many k-mers appear only rarely in the genome. We considered k-mers to be significantly enriched in the test set relative to background if their false discovery rates were ≤ 0.1 (**Supplementary Table 8**).

Cloning, mutagenesis, cell culture and transfection. The human β-globin Dup33 minigene plasmid[59] (gift of D. Black) was used to create splicing reporters. The *Vldlr*(84) and *Nfix*(123) exons and associated intron sequences (~200 bp on each side) were amplified from mouse genomic DNA and inserted between the ApaI and BglII sites. Mutagenesis was performed using the QuikChange Site-Directed Mutagenesis Kit (Stratagene Corp.)

We immortalized mouse embryonic fibroblasts from *Mbnl1*$^{\Delta E3/\Delta E3}$ (C57BL/6J background) by transfection with pOT, a plasmid expressing SV40 T-antigen[60] (gift of X.-D. Fu). We grew the cells in antibiotic-free DMEM with 15% (v/v) FBS. To test the function of Mbnl1 and the motifs, 0.5 μg wild-type or mutant minigenes and 1 μg of pEGFP-MBNL1 (ref. 28) or of the control plasmid producing only GFP (pEGFPC1) were co-transfected using Lipofectamine 2000 reagent (Invitrogen). We harvested RNA 24 h after transfection.

Detection of enrichment of functional gene classes. We evaluated Gene Ontology term enrichment using GoMiner[38] (Version 200904, April 2009; Application Build 246), relative to the set of all genes expressed in the experiment. Because we used Fisher's exact test for scoring, enrichment of GO classes for which there are five or fewer genes must be considered tentative.

57. Tusher, V.G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA* **98**, 5116–5121 (2001).
58. Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408 (2001).
59. Dominski, Z. & Kole, R. Selection of splice sites in pre-mRNAs with short internal exons. *Mol. Cell. Biol.* **11**, 6075–6083 (1991).
60. Hanahan, D., Lane, D., Lipsich, L., Wigler, M. & Botchan, M. Characteristics of an SV40-plasmid recombinant and its movement into and out of the genome of a murine cell. *Cell* **21**, 127–139 (1980).

doi:10.1038/nsmb.1720