Transcriptome-wide Regulation of Pre-mRNA Splicing and mRNA Localization by Muscleblind Proteins

Eric T. Wang,1,2,3 Neal A.L. Cody,5 Sonali Jog,6 Michela Biancolella,6 Thomas T. Wang,1,3 Daniel J. Treacy,1 Shujun Luo,7 Gary P. Schroth,7 David E. Housman,1,2,3 Sita Reddy,6 Eric Lécuyer,5,8 and Christopher B. Burge1,2,3,4,*

1Department of Biology
2Division of Health Sciences and Technology
3Koch Institute for Integrative Cancer Research
4Department of Biological Engineering
Massachusetts Institute of Technology, Cambridge, MA 02142, USA
5Institut de Recherches Cliniques de Montréal (IRCM), Montréal, Québec H2W 1R7, Canada
6University of Southern California, Los Angeles, CA 90033, USA
7Illumina, Inc., Hayward, CA 94545, USA
8Département de Biochimie, Université de Montréal, Québec H3C 3J7, Canada
*Correspondence: cburge@mit.edu
http://dx.doi.org/10.1016/j.cell.2012.06.041

SUMMARY

The muscleblind-like (Mbnl) family of RNA-binding proteins plays important roles in muscle and eye development and in myotonic dystrophy (DM), in which expanded CUG or CCUG repeats functionally deplete Mbnl proteins. We identified transcriptome-wide functional and biophysical targets of Mbnl proteins in brain, heart, muscle, and myoblasts by using RNA-seq and CLIP-seq approaches. This analysis identified several hundred splicing events whose regulation depended on Mbnl function in a pattern indicating functional interchangeability between Mbnl1 and Mbnl2. A nucleotide resolution RNA map associated repression or activation of exon splicing with Mbnl binding near either 3' splice site or near the downstream 5' splice site, respectively. Transcriptomic analysis of subcellular compartments uncovered a global role for Mbnls in regulating localization of mRNAs in both mouse and Drosophila cells, and Mbnl-dependent translation and protein secretion were observed for a subset of mRNAs with Mbnl-dependent localization. These findings hold several new implications for DM pathogenesis.

INTRODUCTION

Muscleblind-like (Mbnl) proteins are a deeply conserved, developmentally regulated family of RNA-binding factors that contribute to muscle and eye development and have been implicated in the genetic disease myotonic dystrophy (DM) (Artero et al., 1998; Begemann et al., 1997; Miller et al., 2000). Flies possess a single gene, mbl, whereas mammals express three closely related Mbnl genes (Fardaei et al., 2002). In mouse and human, Mbnl1 and Mbnl2 are expressed across many tissues, including brain, heart, and muscle, whereas Mbnl3 is expressed primarily in placenta (Kanadia et al., 2003b; Squillace et al., 2002). Mammalian Mbnl proteins contain two pairs of highly conserved zinc fingers, which bind to pre-mRNA to regulate alternative splicing (Pascual et al., 2006). In DM, Mbnls are sequestered away from their normal RNA targets by interaction with expanded CUG or CCUG repeats (Miller et al., 2000). This reverses the normal developmental accumulation of Mbnls, shifting splicing toward fetal isoforms (Lin et al., 2006).

The hypothesis that Mbnl proteins are responsible for a large fraction of the aberrant splicing patterns observed in DM was supported by a splicing microarray analysis (Du et al., 2010), which found that 80% of ~200 alternative isoform changes observed in a CUG-expressing mouse model of DM—including 55 cassette exons—were reproduced in mice lacking functional Mbnl1 protein. These results suggested that CUG-repeat-induced transcriptome changes are largely Mbnl dependent but did not identify direct targets of Mbnl.

Mbnl proteins also have significant cytoplasmic expression and have been proposed to contribute to regulation of messenger RNA (mRNA) stability (Du et al., 2010; Masuda et al., 2012; Osborne et al., 2009) or localization (Adereth et al., 2005). Most mRNAs exhibit specific patterns of localization in the cell, which may be mediated by sequence-specific RNA-binding proteins (RBPs) that interact with cis-regulatory elements located in 3' untranslated regions (UTRs) (Andreassi and Riccio, 2009; Lécuyer et al., 2007; Martin and Ephrussi, 2009). Localized translation of mRNAs is thought to play important roles in proper protein localization and assembly of protein complexes (Besse and Ephrussi, 2008; Lécuyer et al., 2009) and to affect cellular properties such as synaptic plasticity (Horne-Badovinac and Bilder, 2008; Kislauskis et al., 1997; Martin and Zukin, 2006). Localization of mRNAs to the rough
endoplasmic reticulum (ER) via interaction of the signal recognition particle (SRP) with the nascent signal peptide can also involve recognition of sequences in the mRNA to a substantial extent (Pythila et al., 2008). Thus, sequence-specific RBPs like Mbnls that have significant expression in both nucleus and cytoplasm may potentially influence diverse aspects of mRNA biogenesis and function from processing through localization and translation.

To identify direct regulatory targets of Mbnl1, we analyzed the transcriptomes of brain, heart, and muscle tissue isolated from mice lacking functional Mbnl1 protein (Kanadia et al., 2003a) in conjunction with ultraviolet crosslinking/immunoprecipitation/sequencing (CLIP-seq) to identify binding targets in the same tissues and in mouse myoblasts. These data yielded a transcriptome-wide RNA map for Mbnl-dependent splicing regulation. We observed extensive binding to 3′ UTRs of mRNAs encoding membrane and synaptic proteins and uncovered a widespread, conserved role for Mbnls in regulating localization. This function may contribute to the molecular events responsible for DM pathology.

RESULTS

Identification of Hundreds of Mbnl-Responsive Exons

We performed single-end RNA sequencing (RNA-seq) of poly(A)+ RNA from brain, heart, and muscle from five 4-month-old Mbnl1^{TAES/E3} mice and age-matched controls (129/SvJ), yielding 247 million uniquely mapped 48-base reads. Mbnl1^{TAES/E3} mice lack Mbnl with RNA-binding activity (Kanadia et al., 2003a) (Figure S1A available online). MISO was used to compute percent spliced in (PSI or \(J_{BF}\)) for cassette exons that varied in at least one sample relative to control (\(\Delta PSI > 0.1, BF > 5,\) and \(n = 465\) cassette exons). The mean splicing change was only modestly correlated to the level of Mbnl1 (Figure 1D or the level of Mbnl2 (Figure S2A) but showed very strong correlation with total Mbnl levels (Mbnl1 + Mbnl2) (Figure 1E). This trend held not only for exon skipping but also for alternative 3′ splice sites (178 events), alternative 5′ splice sites (107 events), and gene expression changes (Figure S2).

Transcriptome-wide Binding Locations of Mbnl1 in Brain, Heart, Muscle, and Myoblasts

To distinguish between direct and indirect Mbnl targets and to study context-dependent rules for Mbnl regulation, we performed CLIP-seq (with random barcodes in some cases) as described (König et al., 2010; Ule et al., 2005; Wang et al., 2009), yielding 1.6 million, 250,000, 120,000, 71,000, and 2.06 million reads uniquely mapping to the genome and splice junctions in the 129/SvJ brain, C57BL/6 brain, heart, and muscle, and myoblast samples, respectively (after collapsing identical reads). Mapping occurred predominantly to transcribed regions, with enrichment for introns and/or 3′ UTRs (Figure S3).

Mbnl1 CLIP binding locations were consistently observed in different mouse strains, even when comparing different cell types (e.g., reticulon 4, Figure 2A). These sites were essentially uncorrelated with CLIP sites for a different RNA-binding protein, Nova (Zhang et al., 2010), in this example and overall, supporting the specificity of each CLIP data set (Figure 2B). Controlling for effects of pre-mRNA length and gene expression (Yeo et al., 2009), CLIP clusters were enriched for UGC- and GCU-containing 4-mers in myoblasts (Figure 2C) and other tissues. Alternative exons with CLIP clusters within 1 kilobase were more highly conserved than other alternative exons (Figure 2D, controlling for gene expression), supporting their function.

Splicing Regulation by Mbnls Is Dependent on Aggregate Levels of Mbnl Proteins

To supplement our analysis of mouse tissues and to assess the roles of Mbnl1 relative to Mbnl2 in splicing regulation, we stably infected C2C12 mouse myoblasts with lentiviral hairpins against Mbnl1, Mbnl2, or both Mbnls (which did not cause gross changes in morphology or viability) and conducted RNA-seq. Western blotting confirmed reduction in protein levels (Figure S1C) and revealed increased levels of a higher molecular weight isoform of Mbnl2 in Mbnl1-depleted cells and vice versa, suggesting that these factors cross-regulate each other posttranscriptionally.

Expression levels of isoforms were comprehensively assessed (Table S1), and levels of Mbnl1, Mbnl2, and total Mbnl (Mbnl1 + Mbnl2) were determined (Figure 1C). We also calculated the mean splicing change as the mean \(\Delta PSI\) for cassette exons that varied in at least one sample relative to control (\(\Delta PSI > 0.1, BF > 5,\) and \(n = 465\) cassette exons). The mean splicing change was only modestly correlated to the level of Mbnl1 (Figure 1D) or the level of Mbnl2 (Figure S2A) but showed very strong correlation with total Mbnl levels (Mbnl1 + Mbnl2) (Figure 1E). This trend held not only for exon skipping but also for alternative 3′ splice sites (178 events), alternative 5′ splice sites (107 events), and gene expression changes (Figure S2).

Mbnl3 was lowly expressed and did not vary significantly between treatments, so it was omitted from this analysis. Together, these data suggested that Mbnl1 and Mbnl2 function interchangeably in regulation of a large set of splicing targets and that the net depletion of both factors will be relevant in DM and genetic models of the disease.
Context-Dependent Splicing Regulation by Mbnl Proteins

Several individual Mbnl-regulated exons have been well characterized, with Mbnl either activating or repressing exon inclusion (Hino et al., 2007; Ho et al., 2004; Warf and Berglund, 2007; Sen et al., 2010). Binding sites upstream or within the 3’ splice site generally repress splicing (Warf and Berglund, 2007), and enrichment of Mbnl motifs downstream of activated exons suggests that Mbnl may activate splicing in this context (Du et al., 2010; Goers et al., 2010). Our CLIP-seq data confirmed a number of binding sites discovered in previous studies, e.g., sites downstream of Sercal exon 22 (Hino et al., 2007), whose inclusion is highly sensitive to Mbnl levels (Figure 2F). We also uncovered thousands of new locations for Mbnl1 binding associated with splicing activity, including sites in/around exon EIIIB of Fibronectin I (Figure 2E), whose splicing is regulated during development and wound healing (Ffrench-Constant and Hynes, 1989; Ffrench-Constant et al., 1989) and which increased as total Mbnl levels decreased (Figure 1E). Splicing repression was generally associated with

Figure 1. Dependence of Splicing Changes on Total MBNL Levels by RNA-Seq Analysis
(A) RNA-seq read coverage across Trim55 exon 8 from WT and Mbnl1 knockout heart tissue. MISO ψ values and 95% confidence intervals shown at right.
(B) Global correlation of ψ values for cassette exons with dendrogram shown below.
(C) The fraction of MBNL1, MBNL2, or total MBNL remaining in the indicated knockout tissues and knockdown myoblasts relative to control tissues/cells.
(D and E) Mean splicing change (mean |Δψ|) correlates much better with total MBNL depletion than with MBNL1 depletion.
See also Figures S1 and S2 and Table S1.
Figure 2. MBNL Binding Is Conserved and Associated with Context-Dependent Splicing Activity

(A) Mbnl1 and Nova CLIP tags derived from various tissues and cells in the 3' UTR of reticulon 4.

(B) Correlation of CLIP tag densities in five nucleotide windows across all 3' UTRs expressed in brain and C2C12 myoblasts.

(C) Histogram of Z-scores of 4-mers occurring in Mbnl1 CLIP clusters from C2C12 myoblasts, relative to control regions from the same genes.

(D) Alternative exons with Mbnl1 CLIP clusters upstream or downstream are more conserved than other alternative exons expressed at similar levels (shading represents SEM).

(E and F) Dependence on Mbnl levels of ψ values of Fibronectin 1 exon EIIIB and SercaI exon 22, respectively. CLIP data are shown above. MISO ψ values and 95% confidence intervals are shown at right.

(G) Dependence of change in ψ value (WT and Mbnl1/Mbnl2 double knockdown) on Mbnl1 binding in the upstream intron (last 300 bases), alternative exon, or downstream intron (first 300 bases).

See also Figures S3 and S4 and Table S2.
Figure 3. Patterns of MBNL Binding and a Nucleotide Resolution Map of Splicing Activity

(A) Examples of substitutions observed in Mbnl1 CLIP tags and calculation of probability of crosslink-induced substitution (p(CIS)).

(B) Cytosine exhibits the greatest p(CIS) within Mbnl1 CLIP clusters and is most often sequenced as thymine.

(C) Information (relative entropy) of positions adjacent to frequently substituted cytosines as a function of substitution frequency.

(D) Mean CLIP density near exons activated (red) and repressed (blue) by Mbnl1, using binding sites with p(CIS)C\textsubscript{T} > 0.1. Density around exons unaffected by Mbnl depletion is shown by dotted line, with 90% confidence intervals shown in gray. Colored dots denote 100 nucleotide-long windows with a significantly greater density of CLIP sites at regulated than nonregulated exons.
binding in the upstream intron and in the alternative exon, with increased binding associated with stronger repression (Figure 2G). Splicing activation was associated with binding in the downstream intron and also increased with increased binding density (Figure 2G).

**Crosslink-Induced Substitutions Locate Precise Mbnl1 Binding Sites**

To define a high-resolution “RNA map” of the regulatory activity associated with binding at specific distances from splice sites, we sought to pinpoint Mbnl binding sites by analysis of crosslink-induced mutations (Kishore et al., 2011). The frequency of crosslink-induced substitutions (CIS) was calculated at each position as the fraction of reads that cover the position but do not match the reference base (Figure 3A). The base with highest CIS frequency was cytosine, and C to T substitutions occurred at least 5- to 10-fold more often than any other substitution (Figure 3B). Cytosines with high CIS had biases in flanking bases; the –1 base just preceding the substituted C was increasingly biased toward guanine (reaching 94%), and the –2, +1, and +2 bases were preferentially uracil (Figure 3C). The resulting motif closely matched the UGCU motif found in previous studies of Mbnl1 binding affinity and in Figure 2C, suggesting that frequently substituted cytosines represent sites of direct cross-linking to Mbnl1 zinc fingers (Teplova and Patel, 2008). Deletions in CLIP-seq reads were also observed but were rarer and less frequently associated with Mbnl motifs, and so they were not separately analyzed (Figure S3).

The overlap between MBNL-regulated and MBNL-bound exons was quite extensive (Table S2). In all, 31% of regulated exons (compared to 14% of unregulated exons) were associated with high-confidence CLIP sites (P(C → T) > 10%), and this is likely an underestimate, as the fraction increased to 44% in genes expressed above median levels and exceeded 75% when considering all CLIP sites.

We first constructed an RNA map for Mbnl-dependent splicing regulation by using the frequently substituted positions (P(C → T) > 10%, Figure 3D). Activation- and repression-associated binding were ~4.5-fold and ~9-fold above background, respectively (both Z-scores > 10; Experimental Procedures), and occurred in almost completely distinct locations. A repression-associated peak was centered on the alternative exon and extended through the 3′ splice site ~140 bases into the upstream intron, which is consistent with repression via direct occlusion of exonic enhancer, 3′ splice site, or branch site elements. Additional repressive peaks occurred ~100 bases 5′ of the downstream 3′ splice site, suggesting that repression of splicing may occur by different mechanisms (Figure 3D). An activation-associated peak began just 3′ of the regulated exon and extended ~120 bases into the downstream intron. These data establish precise locations at which binding of Mbnl is associated with repression or activation of splicing, enabling prediction of regulatory consequences. An RNA map constructed by using all CLIP sites yielded qualitatively similar results but with peaks that were less consistent and pronounced, suggesting lower precision (Figure S4).

An RNA map constructed for alternative 3′ splice sites revealed significant enrichment of Mbnl binding between the competing splice sites in cases associated with repression of the intron-proximal site by Mbnls (Figure S4D). This pattern of activity is analogous to that observed for other motifs that function as exonic splicing silencers (ESSs) (Wang et al., 2006).

**3′ UTR Binding Is Associated with Specific Protein Functions and Localizations**

Genes with Mbnl binding in 3′ UTRs were most enriched for “cellular component” Gene Ontology categories, as compared to “molecular function” and “biological process” (Figure S4). In all tissues and myoblasts, enriched cellular component categories included synapse, membrane fraction, and insoluble fraction (Figure 3G). Some of these targeted genes included mRNAs encoding proteins involved in synaptic vesicle fusion such as Snap25 and Vamp1, signaling proteins like Camk2α, the extracellular matrix (ECM)-interacting membrane protein integrin β1, and secreted ECM proteins such as collagenes. These data suggested a functional association between Mbnl1 and messages encoding proteins with these localizations. Consistent with this interpretation, the cytoplasmic distribution of Mbnl1 did not appear uniform, with localization at the periphery of the cell or in structures appearing to anchor cellular projections (Figure S4F). Furthermore, as observed previously in human cells (Holt et al., 2009), Mbnl protein partially shifts from the nucleus to cytoplasm following trypsinization and replating of mouse myoblasts (Figure S4E). Genes containing Mbnl-regulated alternative exons were modestly enriched for some of the same Gene Ontology categories as above, but none reached statistical significance (data not shown).

A substantial proportion of Mbnl CLIP reads (sometimes the majority) were located in 3′ UTRs, which is consistent with a recent study (Masuda et al., 2012), and exhibited enrichment relative to RNA-seq reads (Figure S3C). Tetramer motifs in CLIP reads mapping to 3′ UTRs and introns (based on Z-scores, as in Figure 2C) were highly overlapping, suggesting that Mbnl binds similar motifs in 3′ UTRs as in introns (data not shown). Binding occurred predominantly 5′ of the cleavage and polyadenylation site (PAS), which is consistent with interaction in the cytoplasm, and tended to increase closer to the PAS, suggesting a function other than splicing regulation (Figure 3E). Those 3′ UTRs containing CLIP clusters were more highly conserved than other 3′ UTRs of similar length and expression level, suggesting enrichment for functional features (Figure 3F).

**Mbnl1 Preferentially Binds to Localized mRNAs**

Proteins that bind 3′ UTRs are associated with specific cellular and molecular functions, including regulation of mRNA stability, translation, and localization (Moore and Proudfoot, 2009; Zhao

(E) Mean Mbnl1 CLIP density at positions along 3′ UTRs by using three CLIP data sets.
(F) Conservation in 3′ UTRs with and without Mbnl1 CLIP clusters with similar expression levels and UTR lengths (shading represents SEM).
(G) Significant Gene Ontology categories of genes with 3′ UTRs having Mbnl1 CLIP.

See also Figure S4 and Table S2.
Figure 4. mRNA Localization Is Associated with Binding by Mbnl1

(A) Cellular fractionation scheme in C2C12 mouse myoblasts and fly S2R+ cells. Adherent myoblasts were fractionated on tissue culture plates, whereas semiadherent S2R+ cells were fractionated by using centrifugation, yielding “cytosolic,” “membrane,” and “insoluble” fractions.

(B) Western blot of Hsp90, Calnexin, and Tropomyosin I proteins for each compartment in control and Mbnl knockdown myoblasts.

(C) Simplex representation of mRNA expression across compartments in control myoblasts, with colored mRNAs belonging to specific functional categories.

(D) Cumulative distribution of log2(Membrane/Cytosolic) and log2(Insoluble/Cytosolic) values grouped by number of clusters. The number of genes in each cluster is indicated, along with the percentage of clustered genes (M) and genes with no clusters (I).

(E) Cumulative fraction plots for log2(Membrane/Cytosolic) and log2(Insoluble/Cytosolic) values, showing the distribution of gene expression across compartments.

Figure 4 continued...

(F) Continued description of cumulative fraction plots for log2(Membrane/Cytosolic) and log2(Insoluble/Cytosolic) values, showing the distribution of gene expression across compartments.

(G) Continued description of cumulative fraction plots for log2(Membrane/Cytosolic) and log2(Insoluble/Cytosolic) values, showing the distribution of gene expression across compartments.

Figure 4 shows that mRNA localization is associated with binding by Mbnl1. Cellular fractionation schemes are depicted in C2C12 mouse myoblasts and Drosophila S2R+ cells. Western blots of Hsp90, Calnexin, and Tropomyosin I proteins are shown for each compartment in control and Mbnl knockdown myoblasts. Simplex representations of mRNA expression across compartments in control myoblasts, with colored mRNAs belonging to specific functional categories, are also shown.

Cell 150, 710–724, August 17, 2012 ©2012 Elsevier Inc.
et al., 1999). The strong biases for binding to mRNAs associated with particular cellular compartments and a previous study demonstrating that Mbnl2 localizes integrin α3 mRNA to the plasma membrane for translation (Adereth et al., 2005) suggested that Mbnls might play a general role in localization of many mRNAs.

To assess mRNA localization in mouse myoblasts, we used cellular fractionation followed by RNA-seq. Sequential detergent extractions yielded cytosolic, membrane (including organelles such as rough ER), and insoluble fractions (Jagannathan et al., 2011; Vedeler et al., 1991) (Figure 4A). Western analysis of protein markers Hsp90, calnexin, and tropomyosin 1 in each of these fractions confirmed separation of the corresponding cellular compartments (Figure 4B). Hsp90 was exclusively cytosolic (Cserrmely et al., 1998); calnexin, an ER resident protein (Bergeron et al., 1994), was almost exclusively associated with the membrane fraction; and tropomyosin, which decorates the cytosol (Gunning et al., 2005), was substantially enriched in the insoluble fraction. In parallel, we performed similar analysis in Drosophila S2R+ cells to assess potential phylogenetic conservation of effects on mRNA localization. Semiaherent S2R+ cells were processed by sequential homogenization/centrifugation to obtain analogous cellular compartments (Figure 4A). Assessment of several marker proteins in S2R+ cells confirmed successful fractionation (Figure 5B), and the mRNA content of each fraction was analyzed by RNA-seq (Table S3).

To visualize differences in mRNA abundance across the three fractions, we displayed mRNAs as points in a triangular simplex, where the proximity to each corner of the triangle represents the relative expression bias toward the corresponding cellular compartment (cytosolic, membrane, or insoluble). Although this approach does not provide absolute mRNA abundance in each compartment, it allows the comparison of relative mRNA levels across compartments and conditions. mRNAs belonging to particular functional categories showed compartment-biased expression consistent with the localization patterns of their encoded proteins (Figure 4C), which is similar to what has been established for some mRNAs (Besse and Ephrussi, 2008). In myoblasts, mRNAs encoding nucleosomal proteins were enriched in the cytosolic compartment, whereas integral plasma membrane and proteinaceous ECM proteins were biased toward the membrane and especially insoluble cellular compartments relative to the cytosol (Figure 4C). These observations are consistent with previous studies showing that mRNAs encoding secreted proteins are biased toward membranous cellular compartments such as rough ER and nuclear envelope (Pyhtila et al., 2008) and are also consistent with studies showing that mRNAs encoding cytoskeletal or ECM proteins are actively translated on cytoskeletonally anchored ribosomes (Hesketh, 1994; Pachter, 1992; Sundell and Singer, 1991). As expected, mRNAs encoding signal peptides were highly enriched in the membrane and especially in insoluble compartments (Figure S5G). Similar associations between mRNA localization and protein function were observed in Drosophila cells (Figure SSD).

To assess the relationship between Mbnl binding and mRNA localization, we subdivided genes by the number of 3’ UTR CLIP clusters and plotted their relative density within the simplex. We observed that the proportion of mRNAs biased toward the “insoluble” corner increased significantly with the number of CLIP clusters (Figures 4D and 4E). These observations support the association of Mbnl with cytoskeletonally trafficked mRNAs and suggest a connection to the secretory pathway.

**Mbnl Depletion Alters Localization of Bound mRNAs** To assess whether Mbnl directly impacts localization of its binding targets, we performed analogous fractionation/RNA-seq analyses of mouse myoblasts depleted of Mbnl1 and Mbnl2. Comparable separation of cellular compartments in knockdown cells was confirmed by western analysis (Figure 4B), and similar associations between protein function and mRNA localization were observed (Figure SSC), suggesting that depletion of Mbnls did not grossly alter cellular RNA trafficking. Assessment of mRNA abundance in each compartment in control and Mbnl-depleted cells enabled analysis of changes in localization for individual mRNAs. These changes were displayed as vectors in the triangular simplex and were colored to highlight patterns in direction of change following knockdown (Figure 5A). We focused our analyses on the subset of genes that were biased toward the insoluble corner (in which Mbnl binding was most enriched) and further subdivided these genes according to CLIP cluster density. Following Mbnl depletion, these genes tended to have reduced membrane association and increased bias toward the insoluble compartment (Figure 5C). The extent of relocalization correlated with increasing CLIP density (Figure 5C), supporting a model in which Mbnl binding directly affects mRNA trafficking and localization.

Muscleblind proteins are highly conserved across species and have been shown to be functionally interchangeable between mammal and fly in certain cases (Goers et al., 2010). To assess whether the mRNA localization activity of mouse Mbnl1 was conserved to fly, we depleted mbl mRNA from Drosophila S2R+ cells (Figure S5A) and performed fractionation/RNA-seq analysis as above (Figure 5B and Table S4). Again, mbl depletion did not cause gross changes in cellular morphology, viability, or bulk mRNA trafficking (Figure S5D). To assess effects of mbl depletion, we analyzed mRNAs biased toward the insoluble compartment by using UGGU motif density as a proxy for mbl binding (Goers et al., 2010). As in mouse, we observed a bias for relocalization of putative mbl targets away from membrane and toward the insoluble compartment, and, again, the strength of this trend increased with motif density (Figure 5D). Thus, depletion of muscleblind proteins results in similar alterations in localization.
Figure 5. Mbnl Binding Is Associated with Regulation of mRNA Localization in Mouse and Fly and May Contribute to Protein Secretion

(A and B) Change in mRNA localization following depletion of Mbnls in mouse myoblasts and fly S2R+ cells. Arrows represent changes in mRNA localization where the direction change is encoded by color.

(C and D) The aggregate behavior following Mbnl depletion of the subset of genes biased toward the insoluble compartment in control cells is shown in a polar plot in which the radial amplitude is the number of genes that relocalized in the given direction following Mbnl depletion, stratified by density of CLIP clusters (for mouse) or UGCU 4-mers (for fly). The aggregate change in localization toward each compartment is shown in bar plots. p values denote significance when comparing change in localization of MBNL targets versus nontargets (e.g., >2 versus 0–0.5 for mouse myoblasts).
in the localization of putative mRNA-binding targets (Tables S2 and S3) in both mammalian and fly cells.

Mbnls have also been implicated in regulating mRNA stability (Du et al., 2010; Masuda et al., 2012; Osborne et al., 2009). To investigate a potential link between message stability and localization, we analyzed total gene expression of Mbnl 3' UTR-binding targets following Mbnl depletion. Modest increases (<5%) in mean expression of Mbnl targets following Mbnl depletion were observed overall and in mRNAs localized to particular subcellular compartments (Figure S6).

**Mbnl Function Contributes to mRNA Translation**

Because mRNA localization is often coupled to translational regulation, we sought to assess effects of Mbnl at the translational level by performing ribosome footprint profiling (Ingolia et al., 2011) using control C2C12 mouse myoblasts and myoblasts depleted of Mbnl1, Mbnl2, or both Mbnl1 and Mbnl2. These data, representing millions of ribosome-protected mRNA fragments mapping primarily to coding and 5' UTR regions (Figure S6), were used to obtain genome-wide RPKM (ribosome reads per kilobase of mRNA per million mapped reads) measurements that are analogous to RNA-seq-based RPKMs but reflect the ribosomal association of mRNAs rather than mRNA abundance. We observed that subsets of Mbnl 3' UTR targets, in particular those messages that had membrane-biased localization and moved toward the insoluble compartment following Mbnl depletion, had substantially reduced ribosome RPKMs in myoblasts depleted of both Mbnls relative to controls or cells depleted of Mbnl1 or Mbnl2 alone (Figure 5E). By contrast, the overall abundance of these mRNAs (based on RNA-seq) actually increased slightly following depletion of Mbnls (Figure S6), implying that Mbnl depletion results in reduced translation of this subset of mRNAs. Alternatively, these effects could possibly result from changes in mRNA stability dependent on the translational status of the mRNA. The analogous subset of messages biased to the insoluble compartment did not exhibit significantly decreased ribosome footprint RPKMs (Figure 5E).

**Mbnl Function Contributes to Protein Secretion**

Many of the mRNAs biased toward the insoluble compartment that relocalized following Mbnl depletion encode secreted proteins (Figure S5G). Although, classically, recognition of the signal peptide by the SRP targets mRNAs to the secretory pathway (Biology and Dobberstein, 1975a, 1975b), more recent studies have demonstrated that RNA elements at the 5' ends of open reading frames (ORFs) also contribute to targeting of mRNAs to the rough ER or to efficiency of nuclear export (Cenik et al., 2011; Palazzo et al., 2007; Pyhtila et al., 2008). We therefore hypothesized that effects of Mbnl on mRNA localization might play a role in targeting mRNAs to the secretory pathway.

To test this hypothesis, we performed a dual luciferase assay using two efficiently secreted luciferases, Gaussia and Cypridina. We cloned the 3' UTRs of Biglycan and Fibronectin 1 onto the 3' end of Gaussia lucerase; each gene has abundant Mbnl binding in the 3' UTR and substantially changes in localization following Mbnl depletion (Figure 5F). Upon cotransfection with a plasmid encoding 960 CUG repeats, which mimics DM via functional inactivation of Mbnls (Cooper et al., 2009), we observed a 10%–20% reduction in secretion of Gaussia with the Fn1 or Bgn 3' UTR relative to control (Figure 5F). Overexpression of Mbnl1 coding sequence led to a complete rescue of the Fn1 3' UTR-mediated secretion defect and a partial rescue of Bgn 3' UTR-mediated secretion (Figure 5F), suggesting that the effects of CUG repeats on secretion are attributable to the depletion of Mbnls. Taken together, these data and the footprinting results support a model in which Mbnl binding to the 3' UTR of an mRNA promotes productive translation and/or secretion of the encoded protein.

**Mbnl Function Is Associated with Localization of Alternative 3' UTR Isoforms**

Thousands of mammalian genes express mRNA isoforms with different 3' ends as a result of alternative PAS usage, which can be quantitated based on RNA-seq data (Katz et al., 2010; Wang et al., 2008). Alternative 3' UTR isoforms provide natural “reporters” for the function of regulatory elements located in the distal (long isoform-specific) region. We found hundreds of genes whose alternative 3' UTR isoforms had differing localization patterns in myoblasts (Table S4). These longer isoforms tended to be localized toward the insoluble compartment relative to the shorter isoform (Figure 6A). This effect is likely attributable to RNA sequence elements uniquely present in the distal regions, which were frequently bound by Mbnl1 (Figure 6B).

Depletion of Mbnls led to relocalization of specific subsets of Mbnl-bound mRNA isoforms. Focusing on mRNA isoforms biased toward the insoluble compartment as in Figure 5, we observed that Mbnl1-bound long isoforms shifted strongly away from cytosol and toward the insoluble compartment following depletion of Mbnl1 and Mbnl2. More modest shifts were observed for Mbnl1-bound short isoforms (Figure 6C). These effects persisted when restricting the analysis to 3' UTRs whose overall isoform abundance did not change (Figure 57A), suggesting that they resulted from Mbnl activity outside the nucleus rather than from altered PAS choice in the nucleus. This analysis suggests that Mbnl may play a role in regulating isoform-specific subcellular localization of long UTR isoforms, which are particularly abundant in mammalian brain and muscle (Ramsköld et al., 2009).
DISCUSSION

Because of the extreme heterogeneity of clinical symptoms, DM has been described as one of the most variable disorders known in medicine (Harper, 1979). Some of the heterogeneity in DM1 undoubtedly results from the variability in inherited triplet repeat length and from differences in the extent of somatic expansion in different tissues (Ashizawa et al., 1993). This genomic variability interacts with the diverse tissue expression of DMPK mRNA and of Mbnl and CELF family proteins. Thus, even among different patients with the same inherited repeat length, Mbnl proteins may be depleted to different extents in skeletal muscle, brain, or other tissues. A comprehensive understanding of the molecular basis of DM will therefore require not only an understanding of the normal functions of Mbnl proteins in affected cell and tissue types but also an appreciation of the extent to which different functions are impacted by particular degrees of Mbnl depletion.

Our global analyses yielded several insights into Mbnl function and established genomic resources for future work.

Figure 6. Mbnl Regulates Isoform-Specific mRNA Localization via Binding to Alternative 3′ UTRs

(A) The localization of long and short alternative 3′ UTR isoforms is shown at left (as in Figure 4B). Relative enrichment in pairs of cellular compartments is shown for long and short isoforms at right (as in Figure 4E).

(B) CLIP data for the 3′ UTR of Gtpbp4. Localization of the long and short isoforms in normal myoblasts is displayed in solid blue and black circles, respectively, and in myoblasts depleted of Mbnl in outlined blue and black circles, respectively, with arrows indicating change in localization. Bar plot shows change in RPKM (Mbnl-depleted versus control myoblasts) for the long and short 3′ UTR isoforms of Gtpbp4.

(C) Genes were separated into those with and without CLIP clusters in the core or extended 3′ UTR region for short and long isoforms, respectively, and assessed for direction of relocalization following Mbnl depletion. Bar plot shows the extent of relocalization toward each compartment in which the magnitude of relocalization for CLIP targets is normalized relative to nontargets. Error bars represent SEM. See also Figure S7 and Table S5.
functional, modeling, and clinical studies, identifying hundreds of Mbnl-regulated exons and thousands of Mbnl1 binding sites across the transcriptome, many of which are likely to be conserved to human (Figures 2D and 4B). Knowledge of direct regulatory targets should aid in reconstructing the order of events that occur during disease pathogenesis as Mbnl levels are depleted and should provide insights into mechanism.

The transcriptome data indicate that Mbnl1 and Mbnl2 proteins coregulate a set of shared targets and that their splicing functions may be largely interchangeable. The effects of interventions that alter the levels of Mbnl1 or Mbnl2 are therefore likely to depend strongly on the level of the other factor, which will vary substantially depending on the affected tissue. For example, Mbnl2 depletion might produce greater changes than Mbnl1 depletion in neurons and vice versa in muscle, whereas the effects of simultaneous depletion of both factors by CUG repeats should depend more on the size of the total pool of Mbnls than on the level of either factor alone.

We present several lines of evidence for extranuclear functions of Mbnl1 in regulation of mRNA localization and protein expression. A previous study demonstrated that Mbnl2 localizes integrin α3 mRNA to the plasma membrane (Adereth et al., 2005). In this and many other cases, membrane localization is associated with efficient mRNA translation. Here, we show that Mbnls contribute to targeting of hundreds of mRNAs to membranes. Further, we show that the efficient translation of these mRNAs—particularly those that are membrane biased—is dependent on Mbnls. We also observed that Mbnl-bound 3’ UTRs contribute to protein secretion (Figure 5F), perhaps by facilitating targeting to rough ER.

Our results led to a model of the activities of Mbnl proteins in nucleus and cytoplasm that is illustrated in Figure 7. In this model, nuclear Mbnls bind introns or exons to activate or repress splicing, whereas cytoplasmic Mbnls bind 3’ UTRs and facilitate transport along the cytoskeleton to the rough ER or other membranes, presumably through interaction with cellular transport machineries. Nuclear and cytoplasmic functions of Mbnls may be related. Indeed, splicing regulation by Mbnls was associated with 3’ UTR binding; genes with evidence of direct splicing regulation by Mbnl1 were twice as likely to have CLIP clusters in their 3’ UTR (p = 7 × 10⁻⁶, Fisher’s exact test; Figures S7B and S7C). We observed a positive correlation between the density of 3’ UTR CLIP clusters and proximity to Mbnl-regulated exons, but whether this correlation reflects a
functional relationship or simply results from the correlation between high C + G content and short intron length is not clear (Figure S7D).

Both cytoplasmic and nuclear pools of Mbnl are affected in DM (Miller et al., 2000), suggesting that functions of Mbnl outside the nucleus may play a role in DM pathogenesis. Cellular structures particularly sensitive to alterations in mRNA localization, such as synapses and neuromuscular junctions (NMJs), may be perturbed in DM as a result of Mbnl sequestration. Indeed, NMJ degeneration has been observed in diaphragm muscles in mouse models of DM (Panaite et al., 2008). Additionally, motor neurons derived from human DM embryonic stem cells fail to form mature neuromuscular connections with normal muscle (Marteyn et al., 2011), and worms deficient in neuronal muscle-blind expression fail to form distal NMJs, implicating presynaptic dysregulation (Spilker et al., 2012). Other relevant secreted factors include Fibronectin 1 and Thymosin-β 4 (Figure 6E and Table S2). Our findings suggest that downstream consequences of Mbnl’s cytoplasmic functions, such as altered levels of extracellular proteins or circulating factors, may play important roles in disease pathogenesis and could serve as easily accessible diagnostics for monitoring disease progression and assessing response to therapy.

EXPERIMENTAL PROCEDURES

RNA-Seq Library Preparation and Sequencing
Five 16-week-old 129/SvJ Mbnl1 ΔE3/ΔE3 mice and littermate WT mice were sacrificed for isolation of whole brain, heart, and skeletal muscle (vastus lateralis). Poly(A)+ RNA was prepared for single-end RNA sequencing (50 cycles by Illumina GAII). RNA-seq libraries for C2C12 mouse myoblasts were performed similarly (paired-end 2 × 36 cycles by Illumina GAIIx). For the cellular fractionation experiments, Ribo-Zero columns (Epicerin Biotechnology) were used to reduce ribosomal RNA abundance instead of oligo deoxythymidine (dT) beads, and barcodes were used to multiplex three samples per lane (paired-end 2 × 36 cycles by Illumina HiSeq 2000).

CLIP-Seq Library Preparation and Sequencing
CLIP was performed by using 254 nm UV irradiation as previously described (Wang et al., 2009) using whole brain, heart, and muscle or C2C12 myoblasts. Minor modifications to the protocol can be found in Extended Experimental Procedures. Immunoprecipitation was performed by using a polyclonal antibody that does not exhibit cross-reactivity with Mbnl2 (T. Cooper, personal communication).

Cellular Fractionation for C2C12 Mouse Myoblasts
Extraction of cellular compartments from C2C12 monolayers was performed by using cytosolic lysis buffer with 0.015% digitonin, membrane lysis buffer with 0.5% Triton X-100, and cytoskeletal lysis buffer with 1 M NaCl. Cells were washed with PBS between each extraction. Protein and RNA from each compartment were prepared by heating in SDS page buffer or by guanidinium isothiocyanate, phenol/chloroform, and RNeasy mini columns, respectively.

Cellular Fractionation for Fly S2R+ Cells
S2R+ cells were fractionated by using hypotonic lysis and ultracentrifugation. Cells were lysed in hypotonic buffer (20 mM Tris HCl [pH 7.5], 20 mM NaCl, 5 mM EGTA, 1 mM EDTA, 1 mM dithiothreitol [DTT], 1 mM PMSF, 0.15 units/ml aprotinin, 20 μM leupeptin, and 40 units/ml RNase-Out [Invitrogen]), homogenized by Dounce, and spun for 10 min at 1,000 g at 4 °C to clear nuclear and cell debris. The lysate was spun to obtain the cytosolic fraction as the supernatant. The remaining pellet was resuspended in hypotonic buffer containing 1% Triton X-100, homogenized, and spun again to obtain a membrane fraction as the supernatant and an insoluble fraction as the pellet. RNA was extracted by using Trizol (Invitrogen).

Ribosome Footprinting
Ribosome footprint profiling of C2C12 mouse myoblasts was performed as previously described (Ingolia et al., 2011), with several modifications. Instead of isolating ribosomes from polysome gradients, ribosomes were isolated by pelleting through a sucrose cushion. Further details can be found in Extended Experimental Procedures.

Immunoblotting
Bis-tris SDS-PAGE gels (Invitrogen) were used to separate proteins by molecular weight, and protein was transferred to nitrocellulose by using the iBlot transfer apparatus (Invitrogen). Nitrocellulose blots were blocked in 5% milk–PBS with 0.1% Tween-20. Mouse antibodies against Mbnl1 and Mbnl2, MB1a and MB2a, respectively, were kind gifts from the Wolfson Center for Inherited Neuromuscular Disease and were used at 1:300 dilutions. Rabbit Upf1 antibody (Bethyl Laboratories) was used at 1:10 K. Antibodies against Hsp90 (mouse), Calnexin (rabbit), and Tropomyosin I (mouse) were used at dilutions of 1:1K. SuperSignal West Femto ECL reagent (ThermoScientific) was used to detect activity of horseradish peroxidase (HRP)-conjugated anti-mouse and anti-rabbit secondary antibodies.

Cell Culture, Cloning, and Luciferase Assays
C2C12 mouse myoblasts were cultured in high-glucose Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 20% fetal bovine serum (FBS) and penicillin/streptomycin at 37°C and 5% CO2. Knockdowns of Mbnl1 and Mbnl2 in mouse myoblasts were performed via polybrene-assisted infection of lentivirus derived from the pLKO.1 vector, followed by puromycin selection starting 24 hr after infection. RNA and/or protein were isolated from C2C12 mouse myoblasts 7 to 10 days after infection.

Fly S2R+ cells were cultured in 1x Schneider’s Drosophila media (Lonza) supplemented with 10% FBS, 50 units/ml penicillin, and 50 μg/ml streptomycin at room temperature. Double-stranded RNA (dsRNA) synthesis and knockdown in S2R+ cells were performed as described previously (Clemens et al., 2000). dsRNAs were used to knock down cells on days 1, 3, and 4, and cells were harvested on day 5.

The infusion system (Clontech) was used to clone fragments of Bgn and Fn1 into the 3’ of Gausia luciferase (pCMV-Gluc, New England Biolabs N8081). The Fn1 fragment contained the last intron of Fn1, in addition to the 3’ UTR. Liposome mixtures of 400 ng plasmid encoding Gaussia luciferase, 100 ng plasmid encoding Cypridina luciferase, and either 100 ng of a control (pLKO.1) plasmid, 100 ng of a plasmid containing 960 CUG repeats (DT960), or 100 ng DT960 plus 100 ng of a plasmid encoding the Mbnl1 coding sequence were prepared by using Trans-IT (Mirus). The liposome mixtures were added to C2C12 cells in 12-well plates, and luciferase assays were performed 24 hr later. Luciferase assays (Targeting Systems) were used to assay Gaussia and Cypridina activity. Significance between means was computed by using the Mann-Whitney U test.

Computational Analysis of RNA-Seq, CLIP-Seq, and Ribosome Footprinting Data
RNA-seq, CLIP-seq, and ribosome footprinting data have been submitted to GEO (GSE39911). Computational and statistical methods used to analyze short read data are described in Extended Experimental Procedures.

ACCESSION NUMBERS
The GEO accession number for the RNA-seq, CLIP-seq, and ribosome footprinting data is GSE39911. Additional information can be found at http://genes.mit.edu/burgelab/Supplementary/.

722 Cell 150, 710–724, August 17, 2012 ©2012 Elsevier Inc.
SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2012.06.041.

ACKNOWLEDGMENTS

We thank Tom Cooper for the anti-Mbnl1 antibody used to perform CLIP. We thank Andy Berglund and David Brook for insightful conversations about muscleblind and Phil Sharp and Bob Brown Jr. for their comments and suggestions to perform cellular fractionation experiments. We thank Michael Zuker for suggestions on depleting ribosomal RNAs during ribosome footprint library preparation. We thank the MIT Biomicro Center and the IRCM molecular biology core facility for assistance with deep sequencing and the Koch Institute Swanson Biotechnology Center for microscopy assistance. This work was supported by a Poitras Fellowship (E.T.W.), by equipment grant 0821391 from the National Science Foundation, and by grants from the National Institutes of Health to E.L. and N.C.

E.T.W. and C.B.B. designed the study and wrote the paper with contributions from E.L. and N.C.

Received: February 17, 2012
Revised: April 30, 2012
Accepted: June 20, 2012
Published: August 16, 2012

REFERENCES


