EXPLORING THE FUNCTIONAL CONSERVATION OF MUSCLEBLIND (MBL) PROTEINS

by

Julia C. Oddo

B.S. Molecular, Cellular, and Developmental Biology,
Minor Biochemistry 2011

SUBMITTED TO THE DEPARTMENT OF BIOLOGY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTERS OF SCIENCE
IN
BIOLOGY

AT THE
MASSACHUSETTS INSTITUTE OF
TECHNOLOGY

SEPTEMBER 2015

Signature of Author: Signature redacted
Department of Biology
May 22, 2015

Certified by: Signature redacted
Thesis Supervisor: Chris B. Burge
Professor of Biology and Biological Engineering
CSB Program Director

Certified by: Signature redacted
Thesis Co-Supervisor: Eric T. Wang
Research Fellow

Accepted by: Signature redacted
Department of Biology Graduate Co-Chair : Michael T. Hemann
Associate Professor of Biology

EXPLORING THE FUNCTIONAL CONSERVATION OF MUSCLEBLIND (MBL) PROTEINS

by
Julia C. Oddo

Submitted to the Department of Biology
on September 8th, 2015 in Partial Fulfillment of the
Requirements for the Degree of Master of Science in
Biology

ABSTRACT

Muscleblind (Mbl) is an evolutionarily conserved family of proteins involved in many aspects of RNA metabolism, including alternative splicing. Disruption of Muscleblind in several animals lends to a variety of defects and disease, including the multi-systemic disorder Myotonic Dystrophy (DM). Though much is known about the involvement of Muscleblind in DM, there is much basic knowledge of the protein's function to be discovered. We approach this problem by exploring the functional conservation of a diverse subset of Muscleblind homologs.

The functions of Muscleblinds from a basal metazoan, *Trichoplax adhaerens*, a primitive chordate, *Ciona intestinalis*, and the model organisms, *Drosophila melanogaster* and *Caenorhabditis elegans* were compared to human Muscleblind-like (MBNL). The zinc finger RNA-binding domains are the most conserved region between homologs, suggesting a conserved role in RNA binding and splicing regulation. To test this, we used splicing reporter assays with validated human MBNL-regulated mini-genes and performed RNA sequencing experiments in mouse embryonic fibroblasts (MEFs). Additionally, we accessed the subcellular localization of the homologs to determine conservation of extra-nuclear functions.

Reporter assays in HeLa cells showed that the homologs can positively and negatively regulate splicing. Our RNA-seq experiments led us to discover hundreds of endogenously regulated splicing events, including the identity of the transcripts, direction of splicing regulation, types of splicing events, and the magnitude of alternate exon inclusion in the spliced mRNAs. Additionally, we identified a spectrum of splicing events, from those uniquely regulated by a single Muscleblind, to events regulated by all Muscleblinds, and, characterized the variation in splicing activity that exists between homologs. A subset of events regulated by mammalian Muscleblind were oppositely regulated by non-mammalian homologs. Muscleblinds show nuclear-cytoplasmic localization, which suggests conservation in extra-nuclear functions. In conjunction with exon and intron sequences, this information provides a future tool to discover conserved and novel RNA regulatory elements used by diverse Muscleblinds to regulate splicing and in putative cytoplasmic functions. These data could also be used to determine functionally important residues in Muscleblind proteins and help us better understand the protein family.

Thesis Supervisor: Christopher B. Burge, Professor of Biology and Biological Engineering
Thesis Co-Supervisor: Eric T. Wang, Research Fellow

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

INTRODUCTION

Splicing is a complex process that generates protein diversity

Splicing is a co- or post- transcriptional process in which the spliceosome catalyzes excision of introns, or non-coding regions, from a precursor RNA transcript while concomitantly joining exons, or coding regions. This mechanism enables the generation of multiple mRNA isoforms from a single gene, which can lead to multiple protein isoforms. Splicing can be constitutive or regulated; the type of splicing is usually influenced by the ease of recognition and recruitment of the spliceosome to the splice sites in the target RNA. Important *cis* elements including the 5' (donor), 3'(acceptor), and branch sites are used by the spliceosome in two consecutive transesterification reactions to catalyze splicing.

Alternative splicing, or the regulated inclusion of exons, is a process that contributes to the vast diversity observed in fungi, plant, and animal proteomes. Transcripts that are alternatively spliced, as opposed to constitutively spliced, contain suboptimal splice sites which are inefficiently recognized by the spliceosome. There are different types of alternative splicing events including skipped (cassette) exons, retained introns, mutually exclusive exons, alternative 5' splice sites, and alternative 3' splice sites. Trans-acting protein factors can function as regulators of alternative splicing. In general, these factors interact with specific sequences, or RNA secondary structures, termed splicing regulatory elements (SREs), within the RNA transcript to regulate spliceosome recruitment to or interaction with splice sites. These regulators can function in different splicing event types by enhancing (activators) or repressing (repressors) inclusion of an alternative exon. Depending on the biological context, the same splicing regulator may function as an activator or repressor and can act in a spatially-, temporally-, and/or

developmentally-dependent manner to dictate alternative splicing of target transcripts. *(Reviewed in 1, 2, 53).*

Muscleblind is a conserved protein family with roles in RNA metabolism and disease

Muscleblind (Mbl) is a conserved family of RNA-interacting proteins that regulate many aspects of RNA metabolism, including tissue and developmentally-specifc activation or repression of alternative splicing. Plants, fungi, and bacteria lack any protein that resembles Muscleblind, so it appears that this family is exclusive to metazoans, evolving approximately 800 million years ago *(3)*. Typically, invertebrates encode a single *Mbl* gene, whereas vertebrates encode multiple *Mbl* genes. Paralogs can be differentially expressed in a tissue- or developmental-stage specific manner *(29)*. Humans and other mammals have three Muscleblind-like (*MBNL*) genes, *MBNL1-3. (29, 4). MBNL1* and *MBNL2* are ubiquitously expressed in adult tissue, however *MBNL2* predominately functions and is expressed in the brain *(17, Reviewed in 5). MBNL3* is developmentally regulated and is primarily expressed in placental tissue (4) but has been shown to functions in muscle-cell regeneration and differentiation *(6-8, Reviewed in 5). MBNL* paralogs undergo alternative splicing, which can affect the isoforms' localization and activity. Human *MBNL1*, a gene of interest in this study, contains 10 exons that can give rise to at least 10 different splice isoforms *(4, 9,10)*. In particular, we focus on the 41 kDa isoform of MBNL1, which contain exons 1-4, 6-8, and 10 (9,10).

MBNL1 and *2* play a prominent role in the RNA repeat-expansion disease myotonic dystrophy (DM). In this disease, the sequestration of Muscleblind to toxic ribo-nuclear foci leads to malfunction of MBNL proteins and mis-splicing of several mRNAs. The sequestration and resulting aberrant splicing functions relate directly to many DM symptoms *(43)*. In addition to

known functions in alternative splicing regulation, mammalian Muscleblind is involved in other RNA metabolic processes and gene expression including transcription *(11)*, mRNA stability *(12)*, localization *(13-15)* and microRNA processing *(16)*.

RNA binding by Muscleblind proteins

RNA-binding by Muscleblind proteins occurs through highly conserved tandem CCCH-type zinc finger (ZnF) domains *(18, 29)*. Many studies have strived to identify RNA motifs recognized by MBNL1 and the mechanism by which it binds to transcripts and regulates splicing. Fly MBL and human MBNL1 tend to bind YGCY (Y=C or U) containing RNA motifs *(19, 42)*. CLIP-seq and RNA Bind-n-Seq experiments have identified slightly more specific and sub-optimal motifs, including the 4mers GCUU and UGCU, with MBNL1 binding specificity characterized as YGCY + GCUU *(13, 21)*. The number of GC dinucleotides, the spacing between them, and adjacent sequence can influence MBNL1 binding. In vitro experiments showed that for adjacent sequence, U > C > A > G and having a second GC 1-17 nucleotides away confers enhanced binding *(22)*. A crystal structure of MBNL1 ZnF domain bound to CGCUGU RNA shows that it can interact with single-stranded RNA via specific Watson-Crick base pairing with the GC dinucleotide by looping around the RNA. In this model, it is optimal for there to be distance between GC dinucleotides to allow for MBNL ZnFs to bind the GC Watson-Crick face *(23)*. Other studies showed that MBNL1 can bind to paired GCs or structured RNA *(24, 25)*. That MBNL1 can interact with different YGCY arrangements suggests that the protein can adopt different conformations and interact with the RNA in different ways. The location of MBNL binding sites in pre-mRNA relative to a regulated exon impacts its direction of splicing regulation; upstream-binding of MBNL tends to inhibit while downstream binding of MBNL

activates exon inclusion *(25, 14, 42, 13)*.

Muscleblind has important functions in non-human organisms

Conserved alternative exons with MBNL1 binding motifs have been identified in species which diverged between <30 million to >300 million years ago, including mouse, rat, rhesus macaque, cow, and chicken *(26)*. This finding suggests a conserved role for Muscleblind proteins in divergent animals. Important functions for Muscleblind proteins in non-mammalian organisms have been shown. The first Muscleblind was identified in *Drosophila melanogaster,* which contains a single *Muscleblind* gene that can give rise to several splice isoforms and encodes proteins with one or two tandem ZnFs *(18, 27, 28)*. Our study focuses on MBL isoform D (also known as MBL isoform C), which is considered the isoform with most ancestral function *(29)*. *Drosophila Mbl* is expressed in many muscle types, including the developing eye and the central nervous system and has functions in muscle development and photoreceptor differentiation *(30)* When *Mbl* is disrupted in fly through loss of function mutants or in DM fly models expressing toxic RNA, splicing defects occur, yielding eye and muscle phenotypes similar to DM symptoms *(31-34, 59)*. Fly and human Muscleblind proteins have been previously defined as orthologs *(20, 29, 35)*.

Caenorhabditis elegans also has one *Muscleblind* gene, *K02H8.1*, that can give rise to at least six major isoforms with zero or two ZnFs *(36, 37, 29)*. Our study looks at the function of the MBL-1A protein. Worm isoforms have been found in both adult and larval tissue *(37)* and expression analyses reveal MBL in excretory cells, neurons, and spermatheca *(36, 38)*. Disruption of worm *Mbl* shows defects in adult muscle tissue *(37)* and neuromuscular junction formation in motor neurons *(38)*. *C. elegans* DM models show irregular muscle cells, reduced

coordination, motility and lifespan *(39)* that can be partially rescued by *Mbl* expression. Similar DM-related CUG and CCUG repeat-containing toxic RNAs that are bound by human MBNL can be bound by ZnF-containing worm MBL isoforms *(36, 38)*, which suggests some functional interchangeability between human and worm Muscleblind proteins. Molecular mechanisms involving splicing regulation by *C. elegans* MBL have not been directly shown.

Functional conservation of Muscleblind proteins in evolutionarily distant homologs

The function of Muscleblind proteins in many other organisms is unknown. *Ciona intestinalis* is an ascidian (sea squirt) used as a model system for studying the origins of chordates. It is believed to have diverged at least 520 million years ago from its most common ancestor to chordates, allowing for over a billion years of independent evolution from humans *(40)*. *Trichoplax adhaerens* belongs to the phylum Placozoa and is considered one of the simplest free-living animals, representing a primitive Metazoan. Although the evolutionary position of *Trichoplax* is disputed, it is thought that it diverged from other animal phyla in the Precambrian era, at least 540 million years ago *(41)*. Homologous sequences resembling Muscleblind proteins exist in both organisms *(29)* but whether they share molecular functions similar to MBNL/MBL proteins in human, fly, and worm is unknown.

We analyzed sequence conservation and explored the splicing regulatory capacity of MBL proteins from *Ciona*, *Trichoplax*, and *Caenorhabditis* alongside human MBNL1 and *Drosophila* MBL. Using validated splicing reporter mini-genes in HeLa cells over-expressing Muscleblind protein, we found that these homologs can regulate splicing of specific, exogenous pre-mRNAs. RNA sequencing experiments accessing global splicing regulation in mouse embryonic fibroblasts (MEFs) stably expressing Muscleblind proteins showed that the homologs

also regulate hundreds of endogenous targets. Comparing human and non-human Muscleblind proteins showed interesting similarities and differences in their splicing regulatory activity. Muscleblind proteins from human, fly, worm, *Ciona*, and *Trichoplax* are present in the nucleus and cytoplasm. Cytoplasmic localization suggests extra-nuclear activity, which would further extend the functional conservation of non-human MBLs beyond that of splicing regulation. Studying distant Muscleblind proteins may provide insight into ancestral or novel functions that carry over to human MBNL proteins.

RESULTS

The zinc-finger RNA-binding domain is the most conserved region in Muscleblind

To gain insight into the functional conservation of Muscleblind proteins, we selected four diverse organisms including a basal metazoan; *Trichoplax adhaerens*, a primitive chordate; *Ciona intestinalis*, and the model organisms; *Drosophila melanogaster* and *Caenorhabditis elegans* to compare to *Homo sapiens*. Using human *MBNL1* as query, we performed BLAST searches to identify homologs in these organisms. Our hits yielded a hypothetical protein in *Trichoplax*, a predicted protein in *Ciona, Muscleblind D* in *Drosophila*, and *Muscleblind-1a* in *Caenorhabditis*. For clarity we refer to the various homologous Muscleblind proteins as HsMBNL1, TaMBL, CiMBL, DmMBL, and CeMBL for human, *Trichoplax, Ciona, Drosophila* and *Caenorhabditis,* respectively. To quantify the extent of overall homolog protein sequence conservation, we performed Smith-Waterman alignments comparing HsMBNL1 to each homolog. Percent identity and similarity were used to determine amino acid matches and residues with similar properties within a local alignment (**Table1**). DmMBL had the highest identity (50%) and similarity (62.8%) scores. TaMBL showed 44.4% identity, 60.3% similarity; CiMBL 34.8% identity, 47.4% similarity; and CeMBL was found to share 31.7% identity and 42.5% similarity compared to HsMBNL1.

The Muscleblind family is distinguishable by the presence of CCCH-type zinc finger (ZnF) RNA-binding domains, which normally occur as a tandem pair. Multiple-species alignments demonstrate that the ZnF domains are the most conserved regions of Muscleblind proteins (**Figure1**). The number of ZnFs, internal spacing between the conserved cysteine and histidine residues, and spacing between tandem ZnF domains can vary. HsMBNL1 contains four ZnFs, where ZnF1 and ZnF2 (ZnF1/2) make up one domain and ZnF3 and ZnF4 (ZnF3/4) make

up the second. The two domains are separated by 107 amino acids. Spacing within HsMBNL1 ZnF1 and ZnF3 follows a $CX_7CX_6CX_3H$ (where X is any residue) pattern (**Figure 1A**, cyan outlined rectangles) and ZnF2 and ZnF4 have $CX_7CX_4CX_3H$. DmMBL and CeMBL studied here, contain two ZnF both with $CX_7CX_6CX_3H$ internal spacing. Local alignments showed that DmZnF1 and DmZnF2 are most similar to HsZnF1,4 with 87.5% and 70.4% identity. CeMBL ZnFs most resemble HsMBNL ZnF1/2 (75%, 74% identity). *Ciona* and *Trichoplax* MBL proteins have four ZnFs. CiZnF1 has $CX_7CX_6CX_3H$ spacing and shares the greatest residue identity with HsZnF3 (70.8%), while CiZnF2-4 have $CX_7CX_6CX_3H$ spacing and are most similar to HsZnF4 (92% identity), and HsZnF1 (~57%). CiMBL has 189 residues separating its tandem ZnF domains. TaZnF1,3 have $CX_7CX_6CX_3H$ spacing and show the greatest similarity to HsZnF1 or HsZnF4 (both with 57% similarity) and HsZnF3 (65%). TaZnF2,4 have $CX_7CX_6CX_3H$ internal spacing and both look most like HsZnF4. The linker region between TaZnF1/2 and TaZnF3/4 is the shortest, with 89 residues. These subtle differences in the RNA-binding regions may confer differences in Muscleblind function but the overall conserved nature of the domains suggest conserved RNA binding and splicing regulatory functions.

Regions outside of the ZnF domains are conserved and have been shown previously to be important for Muscleblind function. The motifs RD/KWL, or LEV box, and KxQL/NGR, which closely flank the first ZnF pair, are involved in nuclear localization for some human MBNL1 isoforms *(3, 9)*. These motifs are recognizable in DmMBL, CeMBL, and TaMBL. The linker region between tandem ZnF pairs have the next highest density of conserved residues outside of the ZnFs. Mutation and truncation analysis in this linker region has demonstrated its importance for human MBNL splicing activity *(10, 45)*. Furthermore, it has been shown that proline-rich motifs in human MBNL1, some of which lie in the linker region, can interact with Src family

kinases and alter their activity *(44)*. We observe many conserved proline residues in this region of the homologous MBLs. Together, these observations suggest that regions outside of the RNA-binding domains may also be important for Muscleblind function, particularly the residues conserved across such diverse Muscleblinds.

Muscleblind homologs regulate splicing of mini-gene reporters

To initially explore the functional conservation of the Muscleblind homologs, we conducted cell-based splicing assays using reporters. All homologs were cloned into a high-expression vector with a N'-terminal HA-tag. Western blot analysis was used to confirm that the Muscleblind proteins were expressed at similarly high levels in HeLa cells (**Figure 2B**). To conduct the splicing assays, we transiently co-transfected a vector that expressed a Muscleblind protein or an empty eGFP-containing vector (mock control) with a mini-gene reporter construct. The splicing reporter constructs represent previously validated HsMBNL1-regulated genes including human cardiac troponin T type 2 (*TNNT2*) *(19)*, mouse nuclear factor I/X (*Nfix*) *(14)*, human *MBNL1* (MBNL1 auto regulates its own transcript) *(47)*, human sarcoplasmic/ endoplasmic reticulum Ca^{2+}-ATPase 1 (*ATP2A1*) *(42,48)*, mouse very-low-density lipoprotein receptor (*Vldlr*) *(14)*, and human insulin receptor (*INSR*) *(49, 45, 46)*. Generally, these constructs contain an abbreviated version of the gene, which includes the alternative exon and flanking intronic and constitutive exon sequences. Splicing regulation was quantified by finding the average exon exclusion or inclusion (inclusion product/ (inclusion + exclusion product)) and the splicing activity relative to HsMBNL1 (difference between Muscleblind-mediated and mock exon inclusion divided by the difference between human MBNL1-mediated and mock exon inclusion).

Previous reporter assays and high throughout sequencing experiments have shown that HsMBNL1 can act as both a splicing activator and repressor, wherein binding of HsMBNL1 downstream of the regulated exon promotes its inclusion and upstream binding of HsMBNL promotes exon inclusion *(14, 42, 13)*. The six above mentioned genes represent three examples of MBNL-mediated splicing repression (*TNNT2, Nfix, Mbnl1*) and three MBNL-mediated

splicing activation (*ATP2A1, Vldlr, INSR*). Previous studies have used similar assays to establish that DmMBL can positively and negatively regulate splicing of a subset of these reporters *(29, 50);* however, specific mRNA splicing targets and the possible modes of splicing regulation (activation or repression) by CeMBL, CiMBL, and TaMBL are unknown. We aimed to study the splicing regulatory functions of these proteins alongside HsMBNL1 and DmMBL.

Splicing of *TNNT2* is a well characterized example where HsMBNL1 represses inclusion of alternative exon 5. A minimal number of canonical YGCY HsMBNL1-binding motifs have been identified in a region within 50 nucleotides of the alternative splice sites *(46)*. When *TNNT2* is spliced in the absence of HsMBNL1 over-expression (Mock), exon 5 is included in approximately 56% of transcripts. We see a robust splicing response in cells over-expressing HsMBNL1, with a reduction of exon inclusion to 24%, a similar result observed by others *(46)*. We hypothesized that exposing the *TNNT2* reporter, derived from human genomic sequence, to HsMBNL1 in HeLa cells would allow for the strongest regulation compared to Muscleblind proteins derived from non-human organisms, given that the system (including the *cis* and *trans* regulatory elements) would be optimized for regulation by a human protein. When the non-human Muscleblind proteins are over-expressed with the *TNNT2* reporter, we see variable degrees of splicing regulation. CiMBL is most similar to HsMBNL1, showing 26% exon 5 inclusion (represented as 74% exclusion). With respect to HsMBNL1, CiMBL shows 94% of human Muscleblind activity. Following CiMBL, DmMBL shifts exon 5 inclusion to 31% (79% HsMBNL1 activity), TaMBL represses exon 5 inclusion to 35% (64% HsMBNL1 activity) and CeMBL shows the least activity, suppressing the alternative exon's inclusion 48% (23% HsMBNL1 activity) (**Figure 2A**, top left). These results demonstrate that all homologs can act as splicing repressors but don't retain the same activity as HsMBNL1.

We were interested in determining whether the trend in splicing repression observed when using the *TNNT2* reporter is similar to other reporters known to have HsMBNL1-mediated splicing repression. Particularly, we were interested in events with a different layout of functional HsMBNL1 binding sites. *Nfix* and *MBNL1* represent two transcripts negatively regulated by HsMBNL1. *Nfix* exon 8 is an example of HsMBNL1-mediated splicing exclusion. HsMBNL1 negatively regulates inclusion of its own exon 5, which is a feedback mechanism used to help modulate its subcellular localization *(47, 9)*. Both reporters have multiple, clustered YGCY binding motifs in the upstream intron and within the alternative exon (*Nfix*), which is in contrast to the smaller number, closer proximity, and closer spaced YGCY motifs near the splice sites in *TNNT2 (46)*. Over-expression of HsMBNL1 reduces inclusion of the *Nfix* reporter's exon 8 more than three-fold; from 68% in mock-transfected cells to 19% in HsMBNL1-expressing cells. In contrast to *TNNT2*, where CiMBL retained the most activity compared to HsMBNL, CiMBL retains the least activity compared to HsMBNL1 (46% activity) while fly DmMBL represses exon 8 inclusion most similarly to HsMBNL1, with 22% inclusion (93% HsMBNL1 activity). Overall, the trend in splicing repression for *Nfix* follows that HsMBNL1 exerts the strongest repression of exon 8 inclusion followed by DmMBL, TaMBL, CeMBL, and CiMBL (**Figure 2A**, middle left). When we tested the splicing of the *MBNL1* reporter in the presence of Muscleblind proteins we saw a range of exon 5 inclusion from 19% (HsMBNL1) to 44% (CeMBL) compared to 67% in Mock-transfected cells. HsMBNL1 demonstrated the strongest repressive effects on alternative exon inclusion followed by DmMBL, TaMBL, CiMBL, and CeMBL (**Figure2A, bottom left**). For all three reporters, where Muscleblind mediates exclusion of the alternative exon, HsMBNL1 showed the strongest regulation and CeMBL was always among the two poorest regulators. These results show that all HsMBNL1 homologs can negatively regulate

splicing of these reporters, but there is variation in the extent of splicing regulation, which is transcript dependent.

Next we asked whether the HsMBNL1 homologs could positively regulate exon inclusion. We utilized three reporters that differ in their YGCY-binding landscape: *ATP2A1*, *Vldlr*, and *INSR*, in which HsMBNL1 normally enhances inclusion of alternative exon 22, exon 16, and exon 11, respectively. *ATP2A1* contains multiple closely spaced YGCY motifs downstream of alternative exon 22 *(47, 42)*. Without HsMBNL1 over-expression, exon 22 was included 15% of the time. Co-transfection of *ATP2A1* reporter with HsMBNL1 caused a drastic increase in inclusion to 79%. Over-expression of the HsMBNL1 homologs also caused a robust increase in inclusion, showing more than three times the amount of inclusion product than in mock transfected cells. Although the homologs showed reduced spicing activation compared to HsMBNL1, all retained at least 60% HsMBNL1 activity and of the homologs, CiMBL was the strongest splicing activator followed by DmMBL, CeMBL, and TaMBL **(Figure 2A, top right)**. *Vldlr* contains fewer intronic YGCY motifs down-stream of exon 16. HsMBNL1 enhanced exon 16 inclusion from 18% (mock) to 59% and DmMBL activated splicing as well as HsMBNL. CeMBL and TaMBL showed 70% HsMBNL1 splicing activity and CiMBL regulated inclusion about half as well as HsMBNL1 **(Figure2A, middle right)**. *INSR* intron 11 contains multiple HsMBNL1 binding sites *(50)* downstream of regulated exon 11. Of all the minigene reporters tested, *INSR* shows the least change in alternative exon inclusion between mock-transected and Muscleblind-transfected cells, shifting exon inclusion from around 82% to 100% in HsMBNL1 and DmMBL over-expressing cells. CiMBL, CeMBL and TaMBL also activated exon inclusion above that seen in mock-transfected cells **(Figure2A, bottom right)**. As with the HsMBNL1-mediated exon repression, we see that all Muscleblind proteins tested can act as splicing

activators to varying extents in a transcript-dependent manner. Taken together, our results support previous data showing that human and fly Muscleblind can act as both activators and repressors of alternative splicing and we show the novel splicing regulation by CeMBL, CiMBL and TaMBL of reporter constructs.

Muscleblind homologs regulate splicing of hundreds of endogenous transcripts

The splicing reporter analysis showed that all HsMBNL1 homologs can regulate splicing of a selected subset of transcripts. To get a global view of the splicing regulation of endogenous transcripts, we performed RNA-sequencing experiments on cells reconstituted with our HsMBNL1 homologs. Using MEFs null for *MBNL1/2*, we generated a total of six cell lines that stably expressed N'-terminally tagged GFP-Muscleblind proteins or GFP alone. To obtain cells that stably integrated Muscleblind coding sequence and control for expression levels of the integrated protein (so that cell lines expressing different Muscleblind homologs would express approximately equal levels of Muscleblind protein) we used FACS to select a subset of GFP-positive cells. We performed paired-end RNA sequencing on ribosomal-RNA depleted cDNA libraries made from each cell line. For clarity the cell lines are labeled hereafter as GFP, Hs, Dm, Ce, Ci, Ta to denote cell lines expressing GFP or the respective organism's Muscleblind protein. Approximately 38 million reads uniquely mapped to the mm9 genome with minimal rRNA contamination. Percent spliced in (PSI, Ψ) was calculated for all non-UTR related splicing types using MISO *(51)*. Spliced exons were selected by considering the Bayes factor, BF, (comparing GFP Ψ and Muscleblind Ψ) and $|\Delta\Psi|$ (absolute change in alternative exon inclusion between GFP and Muscleblind samples). For most analyses BF$\geq$ 5 and $|\Delta\Psi| \geq 0.1$ was considered significant; these represent exons differentially spliced in Muscleblind-expressing cells compared to GFP-expressing control cells (No Muscleblind). HsMBNL is the only mammalian Muscleblind in our study and it is identical to mouse MBNL1. We didn't expect complete rescue of all Muscleblind-related functions because we introduces a specific *MBNL1* isoform; nevertheless, we thought that adding back HsMBNL1 would most similarly recapitulate endogenous mouse MBNL1 function. Alternatively, we though introducing DmMBL, CeMBL,

CiMBL, or TaMBL would have less conserved function because the proteins are not in their native context. For this reason, many of our downstream analyses (**Figures 5, 6**) compared the splicing functions of non-human MBLs relative to HsMBNL1.

We wanted to know whether cell lines expressing different Muscleblind proteins had similar amounts of splicing regulation. To access this, we calculated the total number of significantly spliced exons (BF≥ 5 and $|\Delta\Psi| \geq 0.1$) in each cell line (**Figure 3A**). Human, fly, and worm Muscleblind proteins regulated between 500-600 exons, while CiMBL regulated slightly more (617) and TaMBL regulated fewer exons (368). Of the regulated exons, we found that 58% of HsMBNL1-regulated exons are activating, where HsMBNL1 enhances exon inclusion. In contrast, all other Muscleblind homologs showed a slight biased towards splicing repression, where the regulated exon is excluded. DmMBL showed the strongest biased, with 340 repressed exons (68%) and 202 included exons, while ~56% of regulated exons are repressed by Fly, *Ciona*, and *Trichoplax* MBLs.

We looked at splicing of different types of regulated exons (**Figure 3B**). Five different types of spliced exons were analyzed, including skipped exons (SE, or cassette exon), retained introns (RI), mutually exclusive exons (MXE) and exons spliced with differential 5' (alternative 5' splice site, Alt5ss) or 3' (alternative 3' splice site, Alt3ss) splice sites. Across all Muscleblind proteins, SE were the most common and MXE were the least common type of regulated exon, with HsMBNL1 showing the strongest preference towards regulating SE (62% of events compared to 48% in DmMBL and 53% in Ce-, Ci-, and TaMBL). When considering the proportion of each exon type, regulated by the different Muscleblind homologs, we see that CeMBL, CiMBL, and TaMBL form a group in which SE makes up about 53%, RI makes up

~19%, MXE makes up ~3%, Alt5ss makes up ~8%, and Alt3ss makes up 17%. HsMBNL1 and DmMbl show slightly different proportions of regulated exon types.

We expected there to be some differences in splicing regulation by the Muscleblind homologs given the differences in the ZnF number and spacing. Regions outside the ZnF domains are diverse between the homologs, and if RNA-binding alone is not sufficient for splicing regulation, these unique residues may afford even more room for differential regulation. **Figure 3C** is a five-way venn diagram with different color ellipses representing different sets of spliced exons that are significantly regulated by Muscleblind homologs (BF≥ 5 and $|\Delta\Psi| \geq 0.1$). Cases in which two or more ellipses intersect represent the maximum number of exons regulated by two or more Muscleblind proteins. This diagram demonstrates that there are many exons exclusively regulated by a single Muscleblind homolog; for example, 259 exons are uniquely regulated by HsMBNL1 and 109 are only regulated by TaMBL. For every combination of intersections there are at least 4 splicing events shared between the respective Muscleblind proteins. When considering exons regulated by HsMBNL and one other homolog, we see that HsMBNL1 shared the largest number of splicing events in common with CiMBL (53 events), followed by CeMBL (43), DmMBL (31) and least with TaMBL (14).

In addition to exploring the number of Muscleblind-regulated exons, we were also interested in the magnitude of $\Delta\Psi$. In particular, is the magnitude of $\Delta\Psi$ similar between exons regulated by different Muscleblinds? For this analysis, we considered spliced exons regulated in the different Muscleblind cell lines with BF≥ 5. A $|\Delta\Psi|$ cutoff was not imposed because we were interested in the full range of $\Delta\Psi$s. By excluding this filter there were more exons than what is shown in **3A** (Hs=670, Dm=594, Ce=610, Ci=712, and Ta=434 events). The mean $|\Delta\Psi|$ for exons in HsMBNL1-expressing cells was 0.30 ± 0.18, which was most similar to that observed

in the Dm (0.31 ± 0.16). Slightly lower values were observed in Ce, Ci, and Ta, where $|\Delta\Psi|$= 0.26 ± 0.15, 0.25 ± 0.15, and 0.24 ± 0.15, respectively. When considering exons regulated in different Muscleblind-expressing cell lines, the mean magnitude of $\Delta\Psi$ was not drastically different. Overall, diverse Muscleblind proteins retain the ability to positively and negatively regulate splicing of hundreds of endogenous mammalian transcripts, including five major types of spliced exons, and, on average, the exons regulated by different Muscleblind homologs show similar magnitudes of $\Delta\Psi$.

Non-mammalian Muscleblind homologs rescue many events regulated by HsMBNL1

We observed that all Muscleblind homologs in this study regulate splicing of hundreds of endogenous transcripts. There are a subset of unique and shared splicing events regulated by both a single homolog alone, and by a combination of Muscleblind proteins. Expression of HsMBNL1 in MEFs is the same as rescuing with the equivalent mouse MBNL1 isoform because the two proteins are 100% conserved. We hypothesized that the regulated exons and respective $\Delta\Psi$ values observed when reconstituting the MEFs with human MBNL1 provides a baseline for how Muscleblind proteins function in mammalian fibroblasts. To access how well non-mammalian Muscleblind proteins regulate spicing in MEFs relative to HsMBNL1 we looked specifically at exons significantly regulated by HsMBNL1 (GFP vs Hs Bayes$\geq$5, $|\Delta\Psi| \geq 0.1$). 584 HsMBNL1-regulated exons, including 337 HsMBNL1-activated and 247 HsMBNL1-repressed exons were assessed. Comparison of Hs $\Delta\Psi$ (GFP Ψ - HsMBNL1 Ψ) and the non-Hs $\Delta\Psi$ (GFP Ψ - non-Hs Muscleblind Ψ), showed highly correlated $\Delta\Psi$ values (**Figure 4A**). For the most part, HsMBNL1-mediated exon-inclusion isoforms (quadrant III) or HsMBNL1-mediated exon-exclusion isoforms (quadrant II) were also inclusion or exclusion isoforms in cells expressing the non-mammalian Muscleblinds; generally, when HsMBNL1 acts as a splicing activator or repressor, non-mammalian Muscleblinds act as splicing activators or repressors, respectively.

Opposite splicing regulation occurs when HsMBNL1 normally enhances exon inclusion and a homolog represses exon inclusion (quadrant I) and vice versa (quadrant III). Of the four non-mammalian homologs, DmMBL shows the poorest correlation (R^2 =0.62) and largest number of exons oppositely regulated. Not all exons that appear to be significant and oppositely spliced actually are, including some regulated exons in Dm. This is because, in this analysis, we

only filtered for significance using a BF comparison between GFP and Hs samples and not GFP vs non-Hs samples. We found that mRNA expression from GFP, Dm, Ce, Ci, and Ta correlated well with Hs (**Figure 4B**). GFP, Ce, Ci, and Ta RPKM had the highest correlation with R^2 values of 0.95-0.96, and Dm expression was less correlated, with a bias towards lower expressed genes (R^2=0.93). Dm gene expression looks different from GFP, Ce, Ci, and Ta. The transcripts studied in this analysis are expressed at roughly the same levels.

To access the splicing activity of non-mammalian homologs relative to HsMBNL1, we looked at the fraction activation or repression of HsMBNL1-regulated exons. Fraction activation and repression was calculated by taking the difference of GFP Ψ and non-human Muscleblind Ψ divided by the difference of GFP Ψ and HsMBNL1 Ψ. So that +1 represents 100% HsMBNL1 activity, 0 represents 0% activity, and < 0 represents oppositely regulated exons, we set fraction activation= (non-Hs MBL Ψ- GFP Ψ)/(Hs Ψ- GFP Ψ) and fraction repression= (GFP Ψ- non-Hs MBL Ψ)/ (GFP Ψ- Hs Ψ). The heat-map (**Figure 4C**) shows that the majority of exons that were regulated by HsMBNL1, were also regulated by the non-human homologs; fewer than 5% of exons were not regulated (fraction=0, white events). At least 76% of all HsMBNL1-mediated inclusion and 78% of HsMBNL-mediated exclusion exons were regulated to some degree. When looking at the number of exons that showed at least 50% HsMBNL1-splicing regulation, CiMBL shared the most HsMBNL1-activated exons (45%) followed by CeMBL (43%), DmMBL (37%) and TaMBL (31%). Some repressed exons, 36-44%, were regulated by the non-Hs Muscleblinds at half the activity of HsMBNL1.

There were a subset of exons regulated as well, or better than HsMBNL1 (fraction ≥1, dark red), highlighted in **Figure 4D**. When we also applied a filter to select for BF comparisons of GFP vs non-Hs Muscleblind ≥ 5, the majority of spliced exons were significant. Many exons

showed opposite splicing regulation by non-mammalian Muscleblinds. When filtered for significance, many events were lost, leaving nine exons oppositely-regulated by DmMBL, seven oppositely-regulated exons by CeMBL and CiMBL, and three oppositely-regulated exons for TaMBL. Some of the same exons are oppositely regulated by more than one non-human Muscleblind. In particular, an exon in Transformer-2 Beta, *Tra2b*, is oppositely regulated by both DmMBL and CeMBL (fraction activation $\approx$ -0.9, $\Delta\Psi \approx 0.1$), where the non-mammalian homologs repress and HsMBNL1 activate splicing ($\Delta\Psi = 0.11$). TRA2B (also known as SRFS10) is a serine/arginine (SR) RNA-binding protein that can regulate RNA metabolism *(52)*. Another exon in Host Cell Factor C1 Regulator 1, *Hcfc1r1*, (which affects the subcellular localization of HCFC1), involved in cell cycle control and transcription during infection of herpes simplex virus (GeneCards), is repressed by CeMBL, CiMBL, and TaMBL (fraction activation < -0.9, $|\Delta\Psi| \approx 0.17$ to 0.39) but activated by HsMBNL1 ($|\Delta\Psi| > 0.19$). Ψ values for exons in *Ganab*, *Tra2b*, *Ndufv3*, and *Zfp275* that are oppositely regulated by the non-mammalian Muscleblind proteins are shown in **Figure 5E**. Further analysis of sequence motifs within these pre-mRNAs may provide insight into why we see opposite regulation.

Variable inclusion of exons regulated by all Muscleblinds suggests differences in their splicing activity

We were interested in looking at exons regulated in all our Muscleblind cell lines; splicing events significantly regulated by a diverse and evolutionary distant set of Muscleblind proteins may provide information on general binding features recognized by all homologs. Thirty-six exons regulated by all Muscleblind proteins were identified, including fifteen HsMBNL1-mediated inclusion exons, and twenty-one HsMBNL1-mediated exclusion exons (**Figure 5A**). Every type of alternative exon, except exons with alternative 5' splice site selection, are represented (**Figure 3B**). As in **Figure 4**, we specifically compared activity of the non-mammalian Muscleblind homologs to HsMBNL1. When the exons were pared down to those regulated by all Muscleblind proteins, we saw that every exon was spliced in the same direction as in Hs; there was no opposite regulation (**Figure 5A,** fraction activation/repression never fell below zero). Overall, we saw much variation in splicing activity between DmMBL, CiMBL, CeMBL, and TaMBL. In some cases, splicing activity was relatively similar between the non-mammalian Muscleblinds; for example, regulation of exons in *Nae1* and *Rnf125*. Other times one homolog greatly exceeded the splicing activity of the others, like exons regulated by DmMBL in the genes *Pbrm1*, *Arrb2*, and *Phf19*. There were also splicing events that were more strongly regulated by all or most non-mammalian Muscleblinds than HsMBNL1 (fraction activation/repression above 1, represented in shades of red), including exons in the genes *Hnrnpk*, *Tjap1*, and *Pphln1*. Interestingly, we saw that two genes (*Cyld* and *Tjap1*) were represented multiple times, suggesting that multiple exons are regulated in the same gene. With the exception of Dm showing a slight bias towards more lowly expressed genes, expression of genes, with exons regulated by all Muscleblind homologs, was similar between cell lines (data not shown).

Seeing variation in splicing activity suggests that the Muscleblind homologs may (i.) interact differently with known HsMBNL1 binding motifs or (ii.) prefer slightly different motifs. We set out to address (i.) by selecting a subset of thirteen SE, which had annotated exon/intron structure, and searching for known HsMBNL1-binding motifs. Looking at sequences in the SE and within 200 nucleotides upstream and downstream of SE, we identified all YGCY and GCTT 4-mers (**Figure 5B, Table 2**). Generally, the sequences show numerous YGCY and GCTT motifs and many instances where these motifs are tightly clustered (*Cyld, Wdr26, Numa1, Git2, Clasp1, Kif3a*). Although not specifically highlighted in **Table 2**, there are many instances where GC di-nucleotides occurred nearby other GCs or YGCY/ GCTT 4-mers. Previous studies have generated an "RNA map" showing patterns of MBNL binding relative to a regulated exon associated with splicing activity and defined binding upstream as repressive and binding downstream as activating *(25, 14, 42, 13)*. Given those findings, we hypothesized that exons regulated by all Muscleblind proteins would follow similar trends in binding/inclusion patterns. We expected to see an enrichment in upstream MBNL1 binding motifs for repressed exons and an enrichment of downstream motifs for included exons. The events that are represented in **5B** are not validated human MBNL1 spliced exons; nevertheless, we hypothesized that these exons would follow similar trends in binding/inclusion patterns. While we didn't see a bias in number of YGCY motifs in the 200 nucleotides flanking either side of the SE, we notice a high occurrence of TGCT and GCTT motifs. About 67% of all 4-mers identified are TGCT/GCTT, where, often times, the two motifs are clustered or occur as TGCTT. These may represent motifs commonly recognized and bound by all Muscleblind proteins studied here.

HsMBNL homologs show conserved sub-cellular localization

Subcellular distribution of a protein can provide insight into its function. To further examine the functional similarities between our Muscleblind homologs, we sought to explore the cellular localization of the proteins. Several groups have investigated the subcellular localization of HsMBNL1 and shown that the 41 KDa isoform is found in nuclear and cytoplasmic compartments, with slight enrichment in the nucleus *(9, 10, 44, 54)*. Similarly, DmMBL has been previously shown to have nuclear-cytoplasmic localization with a predominant nuclear occupancy *(50, 27)* and CeMBL is found in both cell compartments but enriched in the nucleus of *C. elegans* ventral cord neurons *(38)*. Given previous data and our splicing results, we suspected that MBNL1 and all MBLs would show a strong nuclear presence. We were also interested to determine if the proteins show extranuclear expression, particularly CiMBL and TaMBL, in which Muscleblind protein localization has never been studied.

We imaged our MEFs, which stably express approximately equal amounts of GFP-tagged Muscleblind proteins. Cells were plated and fixed 24 hours later for imaging on a high resolution fluorescence microscope. Anti-GFP signal (green) represents MBNL/MBL proteins, the nucleus was labeled with Hoest (blue), and phalloidin (red) was used to stain actin and outline the cells **(Figure 6)**. In agreement with previous data, HsMBNL1, DmMBL, and CeMBL have nuclear-cytoplasmic expression with enrichment in the nucleus. Of all the homologs, CeMBL appears to be most strongly enriched in the nucleus. CiMBL and TaMBL also localization in the nucleus and cytoplasm. Interestingly, CiMBL and TaMBL are slightly concentrated at the cell periphery, potentially co-localizing with actin. TaMBL-expressing cells show strong peri-nuclear staining. Taken together, this demonstrates that all homologs localize to both the cytoplasm and nucleus, however but there may be differences in protein distribution and concentration within the

compartments. Nuclear localization supports our splicing results and a conserved nuclear function. The proteins' cytoplasmic presence provides correlative evidence that there may be conserved function in the cytoplasm.

DISCUSSION

A multi-species alignment between human, fly, worm, *Ciona,* and *Trichoplax* Muscleblind proteins shows that the ZnF-RNA binding regions are most conserved between species. This is not a surprising result given that the ZnF domain is a strongly conserved and defining trait of the Muscleblind family. When comparing the number of ZnFs, internal CCCH residue spacing within an individual ZnF, and identity of each ZnF relative to HsZnFs, there are similarities and differences that may affect function. Fly/worm have one and *Ciona/Trichoplax* have two ZnF pairs. For all organisms, at least one ZnF in each pair has the longer $CX_7CX_6CX_3H$ spacing pattern. Previous studies mutagenizing human MBNL1 ZnF1-4 in all combinations showed that no two ZnF are equal but maintenance of a ZnF pair is important for splicing function. Human ZnF1/2 pair showed higher RNA binding affinity than ZnF3/4 to several known MBNL1 RNA targets *(46)*. In our homologs, CeMBL contains a HsZnF1/2-like pair, while CiMBL and TaMBL retain one ZnF pair that resembles HsZnF3/4 (CCCH residues are conserved throughout all ZnFs, so it is the identity of the 'X' residues within the '$CX_7CX_6CX_3H$' that dictate similarity). Although the paired distribution of ZnFs is maintained, all homologs, except CeMBL, have at least one pair with non-HsZnF-like identity: CiMBL's second ZnF pair is HsZnF3,3-like, DmMBL's only ZnF pair is Hs-ZnF1,4-like, and TaMBL's first ZnF pair is HsZnF1,4 or 4,4-like **(Figure 1)**. Finding that most homologs have different HsZnF-like pair identity but retain paired ZnF architecture corroborates the idea that it is maintenance of a ZnF pair, perhaps for proper domain folding or RNA interactions, that is important for function.

Assays using splicing reporters showed that our Muscleblind homologs can activate and repress splicing of several human transcripts to varying degrees. In most cases, HsMBNL1 and DmMBL function as the strongest splicing regulators, but there is no specific order for how well the different MBLs regulate splicing. Instead, it appears that the strength of splicing regulation is transcript dependent.

Variation in splicing regulatory activity could arise from differences in protein sequence/ structure. A previous study showed that binding of human MBNL1 to RNA substrates (the RNA substrates have sequences corresponding to the reporter constructs used in this study) generally correlates well with splicing activity and a greater number of intact ZnF pairs enables MBNL1 to bind RNA with higher affinity *(46)*. We saw no correlation between ZnF number and splicing activity; there isn't a preference for Muscleblind proteins containing four ZnFs to regulate reporter splicing better than those with two. Purcell et al., generated human MBNL1 RNA-interaction (RIM) mutants in which two of the four ZnFs have been rendered non-functional. These RIMs included mutants with two functional ZnFs that resemble DmMBL or CeMBL ZnFs. DmMBL ZnFs are HsZnF1,4-like and CeMBL ZnFs are HsZnF1/2-like, which correspond to the human MBNL1 2,3RIM and 3,4RIM mutants, respectively *(46)*. In 2,3RIMs, the functional ZnFs are not in a paired configuration (only ZnFs 1 and 4 are functional) thus RNA binding and splicing activity is poor compared to 3,4RIMS, which have a functional ZnF1/2 pair. Although DmMBL ZnFs have similar identity to the functional 2,3RIM ZnFs, it regulates splicing of our reporters better than what was shown for the 2,3RIMs. This is likely because ZnF pairing is maintained in DmMBL. This further supports the idea that it is the paired-ZnF layout rather than the specific non-CCCH residue identity of the ZnFs that affects function.

Perhaps a more interesting comparison is between CeMBL and the 3,4RIM mutant because, in this case, both proteins maintain a functional ZnF pair. Splicing activity for both CeMBL and 3,4RIM is generally high, but for some reporters, one protein or the other acts as a better regulator. For example, the splicing of *Vldlr* by 3,4RIM had a measured activity of 33% *(46)*, but in our study, we observed 71% activity by CeMBL. For splicing of the *TNNT2* reporter, the activities are flipped; 3,4RIM regulates better than CeMBL. Both studies followed similar splicing assays and used the same reporters. Because the number and identity of the ZnFs is essentially the same in these two proteins, non-ZnF residues likely contribute, to some extent, to the differences in activity. Our multi-species alignment (**Figure 1B**) shows conserved and unique residues outside the ZnF domains, particularly in the linker region between ZnFs. Previous studies have shown that these regions can have important functions in RNA binding and splicing activity *(3, 9, 10, 44, 45)*. It could be that these regions lend to RNA target specificity or protein-protein interactions that affect splicing regulatory activity.

Protein homology can be defined as shared ancestry, which is often identified through sequence conservation. A protein is considered orthologous relative to another if it fits two main criteria: Firstly, the proteins belong to different species and arose from a common ancestral gene via speciation. Secondly, the orthologous proteins maintain similar biological function over the course of evolution. All Muscleblind proteins studied here came from different species belonging to Metazoa and are thought to have arisen 800 MYA (3). Previous research has already demonstrated orthology between DmMBL, CeMBL and HsMBNL1. Our reporter splicing results support that CiMBL and TaMb, used in this study, are also HsMBNL1 orthologs due to their conserved ability to regulate splicing.

To get a more global view of how the Muscleblind orthologs regulate splicing of endogenous transcripts, we performed RNA-sequencing experiments. We found over 350 exons regulated by each Muscleblind ortholog, including uniquely regulated exons and those regulated by multiple orthologs. Over thirty exons are regulated by all orthologs. Both included and repressed exons are represented, but there is a slight bias towards repressed exons in non-mammalian MBLs, with DmMBL showing the strongest bias (**Figure 3**). Generally, the direction of splicing regulation is dictated by where alternative splicing regulators bind splicing regulatory elements (SREs) in the mRNA relative to a regulated exon. Binding to these sites is thought to modulate spliceosomal splice site selection. SREs include intron and exon splice site enhancers (ISEs, ESE), which facilitate exon inclusion (splicing activation), and splice site suppressors (ISS, ESS), which enhance exon exclusion *(Reviewed in 53)* . One interpretation of why there is a bias towards repressed exons is that there is a larger proportion of transcripts with Muscleblind-specific ISSs and ESSs than ISEs and ESEs. If Muscleblind has more opportunity to bind repression-associated sites it may act as a splicing repressor more often than a splicing activator. MBNL1 cross-linking immunoprecipitation (CLIP) experiments in mouse tissues showed that there is more above-background binding of MBNL1 to repression-associated sites (~9-fold) than there is to activation-associated sites (~4.5 fold) *(13)*. A larger proportion of repression-associated binding CLIPs suggests there may be more potential for Muscleblinds to act as repressors. Although HsMBNL1 showed a slight bias towards splicing activation, non-HsMBLs show a bias towards splicing repression. It could be that these proteins bind better to the identified repressive CLIP targets. Motif analysis will need to be performed in order to determine putative binding motif preferences for the different orthologs.

Not only can the orthologs regulate splicing in both activating and repressive directions, they also regulate five major types of alternatively spliced exons (**Figure 3B**). For all Muscleblinds, skipped exons are the most common type of splicing event. When one considers the direction of splicing regulation and proportion of each type of alternative exon, three groups are formed. The groups consist of Hs alone, Dm alone, or Ce, Ci, Ta, where the number of included and repressed exons regulated in Hs and Dm differ from each other and Ce/Ci/Ta, while the number of included or repressed exons in Ce, Ci, and Ta are more similar to one another. Similarly, the proportion of alternative exon types regulated in Hs and Dm is different from the proportion of regulated exon types seen in Ce,Ci, and Ta. Whether this implies that CeMBL, CiMBL, and TaMBL are more functionally conserved than Hs and Dm requires further investigation. Overall, there are a large number and diverse types of Muscleblind- activated and repressed exons.

We found that some exons are uniquely regulated by a single Muscleblind **(Figure 3C).** This suggest that there is something unique about the *cis* element(s) or the *trans* protein(s) influencing the splicing events. The splicing event may be defined by *cis* SREs that are preferentially recognized by only one Muscleblind. It is also conceivable that differences in Muscleblind protein sequence/folding cause different modes of protein-RNA interactions that allow unique regulation to occur. Further motif analyses of the uniquely regulated splicing events and directed mutagenesis of the Muscleblind proteins would need to be preformed to begin to understand this phenomenon.

We specifically looked at spliced exons regulated by HsMBNL1 and asked whether non-mammalian MBLs could regulate them (**Figure 4**). In general, we found that many exons

regulated by HsMBNL1 are regulated, to some extent, by non-mammalian orthologs. The exons are almost always regulated in the same direction. Furthermore, there are exons that are either not regulated, regulated better, or regulated oppositely by the non-mammalian MBLs. Again, these differences may be due to the nature of *cis* SRE recognition. For cases in which there is no regulation, it would make sense that the particular MBL doesn't or very poorly recognizes HsMBNL1 SREs in the context of the transcript. On the other hand, exons that are regulated more strongly by the orthologs may have SREs that are recognized better by the non-mammalian MBLs than by HsMBNL1.

Interestingly, we identified some significant exons that are oppositely regulated by HsMBNL1 and non-mammalian MBLs. Other incidences of closely-related splicing factors showing opposite splicing regulation exist. For example, the cell-type specific alternative splicing regulators PTB and nPTB oppositely regulate exons in *Rip3* and *Exoc1*, where PTB represses exon inclusion and nPTB enhances it *(63)*. In our study, we found that some exons, like exons in *Tra2b* and *Hcfc1r1*, are oppositely regulated by more than one non-mammalian Muscleblind, suggesting a potentially conserved role for this opposite regulation. Muscleblind proteins showing opposite regulation would need to bind the opposite side of a regulated exon compared to where HsMBNL1 binds to uphold the "RNA map" defined for MBNL1 *(13)*. If all Muscleblinds are exposed to the same *cis* elements within the pre-mRNA, opposite regulation would occur if the SRE(s), used by HsMBNL1 to mediate its regulation, wasn't preferentially used by the opposite-MBL regulator. This could occur if the SRE isn't an optimal binding motif for the particular MBL regulator, or, if another, more preferential, site on the opposite side of the regulated exon was present.

Splicing regulatory domains, which are distinct from the RNA-binding domains, have been identified in human MBNL1 protein *(45)*. Multiple studies have suggested mechanisms in which MBNL1 partakes in protein-protein interactions with spliceosome components or cofactors to facilitate splicing regulation *(45, 46, 59)*. Given these insights, another plausible explanation for opposite regulation is that HsMBNL1 and the oppositely-regulating ortholog differentially interact with *trans* factors that are required to regulate the alternative exon.

Over thirty exons are significantly regulated by all Muscleblinds in this study **(Figure 3 and 5)**. Five major types of alternative exons, except Alt5ss **(Figure 3B)**, are regulated and the direction of regulation is always the same for mammalian and non-mammalian Muscleblinds **(Figure 5A)**. For these exons, we observed much variation in the degree of inclusion, suggesting different splicing activity between the non-mammalian MBLs compared to HsMBNL1. As previously discussed, this variation could arise due to differences in recognition of *cis* SREs or interactions with *trans* regulatory factors, which may arise due to differences in the Muscleblind protein sequence/structure. Further inspection of *cis* SREs, showed a plethora of known MBNL binding motifs (YGCY and GCUU) and an enrichment for T(U)-rich GCs, which sometimes occur in a highly clustered arrangement. The motif identities, layout, and/or potential secondary structure present in pre-mRNAs regulated by our Muscleblinds may represent conserve elements used by all Muscleblinds. Motif and RNA structure analysis could be used to help determine if this is the case.

Extra-nuclear functions of human, fly and worm Muscleblind proteins have been described. The Muscleblind family has been implicated in regulating mRNA export/ localization, mRNA decay, and synapse/neuromuscular junction (NMJ) formation. Mammalian MBNL2 can

localize integrin-α3 to the plasma membrane, where its proper localization is important for mRNA translation *(15)*. MBNL binding sites have been identified in the 3'UTR of mRNA and MBNL binding is implicated in affecting the mRNAs cellular distribution and stability *(12, 13)*. Synapse and NMJ defects are observed in *Mbl 1*-deficient worms and DM mouse models *(38, 60)*. We wanted to see if extra-nuclear Muscleblind functions are conserved in the proteins studied here. We found that, like mammalian HsMBNL1, MBL orthologs localize to nuclear and cytoplasmic cell compartments. For human, fly, and worm, this is in agreement with other studies *(9, 10, 27, 38, 44, 50, 54)*. *Ciona* and *Trichoplax* MBLs have some interesting differences in cellular compartment distribution. The presence of these proteins outside the nucleus provides correlative evidence for cytoplasmic functions and the unique localization patterns observed in *Ciona* and *Trichoplax* may highlight differences in cytoplasmic functions. So far, we can only conclude that there is conserved localization; further experimentation needs to be done in order to determine if any of the above-mentioned cytoplasmic Muscleblind functions are retained.

Given that *Trichoplax* is a very simple organism, with only four described cell types, and no evident sensory, muscle, or nerve cells *(41)*, the function of TaMBL, particularly in the cytoplasm, is very intriguing. *Ciona intestinalis* is a basal chordate. This phylogenetic placement is based strongly on morphological/developmental characteristics shared between *Ciona* and other vertebrates. However, many of *Ciona's* morphological similarities to vertebrates are lost during metamorphosis from its tadpole-like larvae to adult stage *(61)*; adult organisms look only vaguely familiar to other vertebrates. *Ciona* underwent gene loss, resulting in genome reduction, with an estimated loss of 45% more ancestral gene families than humans *(62);* interestingly, it

retained a gene encoding a Muscleblind protein. Other Muscleblind functions in these organisms is left to be explored and our data provides a useful tool to further do so.

The data presented here supports conservation in Muscleblind splicing regulation and protein subcellular localization. This data can lend insights into conserved and novel RNA regulatory elements, that may be used by evolutionarily distant Muscleblinds, for various functional purposes. Furthermore, we can use information about the splicing activity and motif preferences in conjunction with protein sequence differences for guided mutagenesis to better define functionally significance residues in Muscleblind proteins.

MATERIALS AND METHODS

Cloning, Cell Lines, and Splicing Reporters

N-terminal HA-tagged DNA constructs encoding HsMBNL1, DmMBL, CeMBL, CiMBL, and TaMBL were cloned into the pCI plasmid (Promega) for use in the splicing reporter assays. HA-HsMBNL1 was PCR-amplified out of pCDNA3 plasmid using primers 6 and 18 (See Table 1: Primer) and inserted into pCI with Xho1 and Not 1 restriction sites. HA-tagged HsMBNL1 in pCDNA3 was previously cloned from MBNL1-eGFP *(19)* obtained from the laboratory of Maury Swanson. HA-DmMBL was cut directly out of a previously cloned construct of HA-DmMBL in pcDNA3 using Kpn1 and Xba1 and inserted into pCI. CeMBL DNA with N-terminal Bam H1 and C-terminal Not 1 restriction enzyme cut sites was synthesized by GenScript and received in pUC57 cloning vector. To add an N-terminal HA-tag to CeMBL, the insert was sub-cloned into HA~pcDNA3 using primers 1 and 2 and Bam H1/Not 1 sites. . Kpn1 and Xba1 restriction enzymes were used to insert Ha-CeMBL into pCI. CiMBL DNA with N-terminal HA-tag and EcoR1 cut site and C-terminal Sal1 restriction enzyme cut sites was synthesized by GenScript and received in a pUC57 cloning vector. HA-CiMBL was PCR amplified and inserted into pCI using EcoR1 and Sal1 HF. TaMBL cDNA was generously provided by Dave Anderson (Thorton Lab, University of Oregon). Primers 23 and 24, which provides N-terminal Xho1 restriction enzyme cut site and C-terminal Bam cut site were used to amplify TaMBL, which was then sub-cloned into pCDNA3 containing an HA insert. Ha-TaMBL was digested out of pCDNA3 and inserted into pCI using Kpn1 and Xba1 restriction enzymes. All primers used for cloning into pCI can be found in Table 1.

N-terminal GFP-tagged constructs encoding HsMBNL1, DmMBL, CeMBL, CiMBL, and TaMBL were cloned into the pUC156 containing PiggyBac Transposon sequences, to generate a vector that was stably introduced into MBNL1/2 null MEFs. These were the cell lines used in our RNAseq and IF experiments. The In-fusion cloning system (Clontech) was used according to the manufacturers instructions with primers 25-33 (**Table 1**) to generate GFP-Muscleblind fusion flanked by PiggyBac Transposon in pUC156 vector. At 60% confluencey, MEF cells null for MBNL1 and MBNL2 were transfected with 2ug plasmid encoding GFP-Muscleblind using TransIT (Mirus) and 500ng of PiggyBac (SBI) transposon to stably introduce GFP-Musclebind into the cells. After 24hrs the cells were subject to puromycin selection (2ug/ml) and sorted for similar GFP-expression. The cell lines were maintained with puromycin.

Reporter constructs used for the splicing assay were previously cloned; TNNT2 was gifted from laboratory of Thomas Cooper *(55, 56)*, ATP2A1 *(42)*, *MBNL1 (47)*, INSR was from Nicholas Webster *(57)*, Nfix, and Vldlr were from Manuel Ares Jr.) were used *(14)*.

Cell Culture

HeLa cells were maintained as a cultured monolayer in Dulbecco's modified Eagle's medium (DMEM) + GLUTAMAX media which was supplemented with 10% Fetal Bovine Serum, FBS, and 10% antibiotic-antimycotic (Gibco, Invitrogen). The cells were kept at a constant temperature of 37°C in a humidified incubator (5% CO_2). MEF cells were maintained in DMEM supplemented with 20% FBS and penicillin/streptomycin at 37°C and 5% CO_2.

Western Blots

Total harvested HeLa cells were suspended in RIPA lysis buffer (100mM Tris pH 7.4, 300mM NaCl, 10% NP40, 10% Na-deoxycholate, protease inhibitor, 200mM PMSF, 10% SDS), and samples went through three freeze/thaw cycles. Protein concentration was quantified using BCA reagent (Thermo Scientific) following manufactures instruction. Total protein lysates (5mg) were loaded on 12% SDS-Page denaturing gel, electrophoresed for 40 min and transferred via a fast partial wet transfer (200mA and 100V for 2 hours) to a 0.45-μm pore size nitrocellulose membrane (GE Water & Process Technologies). Following protein transfer, a ponceau stain (Sigma-Aldrich) was performed in order to ensure proper transfer. TBST was used to wash the nitrocellulose. The blot was blocked for 4 minuets in 4% milk in TBST prior to administration of the primary antibody. All blots were exposed to primary antibody at a dilution of 1:1000 (antibody: 4%milk in TBST) overnight at 4°C and exposed to secondary antibody at a dilution of 1:2000 for 2 hrs at room temperature. HA-probe (F7) mouse polyclonal IgG antibody and actin (I-19) rabbit polyclonal IgG (Santa Cruz Biotech) primary antibodies were used.

Reporter Splicing Assay

HeLa cells were plated in 6-well plates at a density of 1.6-1.8 × 10^6 cells/well. At 80-90% confluencey, the cells were co-transfected with 500ng/well of plasmid containing the splicing reporter and plasmid encoding either a Muscleblind protein or GFP (mock control) using Lipofectamine2000 reagent (Invitrogen) in OPTI-MEM reduced serum medium (Gibco, Invitrogen). After 4 hours incubation in the reduced serum, media was replaced with high-growth media, DMEM+ GLUTAMAX, and the cells were allowed to incubate for 18-24 hours prior to harvesting with TripleE (Gibco, Invitorgen). Experimental procedures follow those previously described *(42)*. RNA was harvested using an RNAEasy Kit (Quiagen) according to the manufacturers instructions. 500ng of RNA from each sample was subjected to a DNAse reaction using RQ1 DNAse (New England Biolabs) following manufacturers instructions. DNAsed RNA [2ul (100ng)] was reverse transcribed (RT) with SuperScript II reverse transcriptase (Invitrogen), according to Invitrogen's protocol, except that half the amount of recommended amount of SuperScript II was used. cDNA was PCR amplified for 20-26 cycles using reporter-gene specific primers. PCR products (5ul) were dyed with Syber Green DNA loading dye (Invitrogen) and resolved on a 6% native acrylamide gel (19:1). The resulting gel was imaged and quantified using AlphaImager and associated software (Alpha Innotech).

Percent exon inclusion was calculated by dividing the background-corrected amount of inclusion splice product by the total amount of splice product (background-corrected inclusion splice product+ background-corrected exclusion product). Splicing activity relative to human MBNL1 was calculated in the following way: (non-HsMbl- mock inclusion)/(HsMBNL- mock inclusion).

RNA-Sequencing

Total RNA was isolated from GFP-Muscleblind expressing MEF cells using Direct-zol RNA columns (Zymo Research) according to manufactures instructions. cDNA libraries were generated starting with 1ug RNA. Briefly, RNA was fragmented, depleted of ribosomal RNA using Ribo-Zero-Gold kit (Epicentre), and reverse transcribed followed by end-repair, adenylation, and adapter ligation. Unique barcodes were used for each library to allow for multiplexing all samples in a single lane (80+80 bases, paired-end, NextSeq). Spliced transcripts

alignment to a reference (STAR) *(58)* was used to map reads to the mouse mm9 genome and mixture-of-isoforms (MISO) *(51)* was used to quantify splicing regulation.

Immunofluorescence

MEF cell lines expressing GFP-Muscleblind were plated (~1.25 x 10^5 cells/well) on collage-coated coverslips (100ug/ul), incubated overnight (~18hr) and then fixed using 4% parafromaldehyde, 15 min at room temperature. Cells were washed in phosphate buffered saline (PBS), permeabilized in 0.2%Triton-X/PBS at room temperature for 3 minuets, blocked in 10% bovine serum albumin (BSA)/PBS at 37° for 30 minuets and exposed to primary antibody, 1:1000 chicken IgY α-GFP (Aves) diluted in PBS + 1% BSA, at 4° overnight. After washing in PBS, the secondary antibody, α-chicken 488 at 1:400 and 594 phalloidin (LifeTechnologies) at 1:400 diluted in PBS + 1% BSA, was incubated on the coverslips at 37° for 1hr. Coverslips were washed in PBS and subject to Hoescht nuclei-stain at room temperature for 10 minuets before final PBS washes and mounting onto microscopic slides. All cells were imaged on an Applied Precision DeltaVision Microscope at 60X magnification, optical sections were deconvoluted using the associated software, and processed using ImageJ. Adjusted intensity projections were generated from the average of three z stacks, centered around the nucleus.

PRIMERS TABLE

	Primer	Sequences	Use
1	ceMBNL3 fwd	5'- CGCGGATCCGCGATGTTTGAT GAAAACAGTAATGCAGCAGGC-3'	Cloning; PCR amplifying ceMBNL out of pUC57 parent plasmid
2	ceMBNL3 rev	5'-TTTATAGCGGCCGCATATTTT TAGAACGGCGGCGGCT-3'	Cloning; PCR amplifying ceMBNL out of pUC57 parent plasmid
3	cTNT RT fwd	5'-GTTCACAACCATCTAAAGCAA GTG-3'	Splicing: PCR amplify minigene RT sample
4	cTNT RT rev	5'-GTTGCATGGCTGGTGCAGG-3'	Splicing: PCR amplify minigene RT sample
5	Dup1 (D1) Rev	5'-GCAGCTCACTCAGTGTGGCA-3'	Splicing: Reverse transcribe (RT) minigene mRNA from cell samples
6	Dup8 Fwd	5'-GACACCATGCATGGTGCACC-3'	
7	*MBNL Exon 4 Fwd	5'-GATCAAGGCTGCCCAATACCAG-3'	PCR Mbnl1 minigene-derived cDNA
8	*MBNL RT Rev	5'-CAGATTCATTTATTAAGAAACCCCAC CCCTTAC-3'	PCR Mbnl1 minigene-derived cDNA
9	HA-Mbl ZnF1-4 fwd	5'-ACGCGTCGACGTCGGATGTAC CCATACGACGTACCAGATTACGCTCT CGAGATGGCCACCGTT-3'	Cloning: PCR amplifying Mbl ZnF 1-4 out of pUC57 parent plasmid and adding N-terminal HA-tag
10	IR RT Rev	5'-GCTGCAATAAACAAGTTCTGC-3'	Splicing: Reverse transcribe (RT) minigene mRNA from cell samples
11	IR Ex. 10 fwd	5'-CGAATTCCGAATGCTGCTCCTG TCCAAAGACAG-3'	PCR IR Minigene-derived cDNA
12	IR Ex. 12 rev	5'-TCGTGGGCACGCTGGTCGAG-3'	PCR IR Minigene-derived cDNA
13	JPp48 fwd	5'-CCGCTCGAGCGGATGGAGTAC CCATACGACGTACCAGATTACGCTAT GGCTGTTAGTGTCACACCAATTCGG G-3'	Cloning; PCR amplifying HA-MBNL out of pcDNA3 parent plasmid and adding an N-terminal EcoR1 cut site
14	MBNL1 RT rev	5'-CAGATTCATTTATTAAGAAAC CCCACCCCTTAC-3'	Splicing: PCR amplify minigene RT sample
15	pcDNA3 F1	5'-ATTAATACGACTCACTATAGG GAGACCC-3'	Cloning
16	pcDNA3 R1	5'-AGCATTTAGGTGACACTATAG AATAGGG-3'	Cloning, Splicing: Reverse transcribe (RT) minigene mRNA from cell samples
17	pCI fwd	5'-GCTAGAGTACTTAATACGACT CACTATAGGC-3'	Cloning; General screening procedure

18	pCI rev	5'-CGCCCATGCAGGTCGAC-3'	Cloning; General screening procedure
19	Serca1 Fwd	5'-ACCTCACCCAGTGGCTCATG-3'	Splicing: PCR amplify minigene RT sample
20	Serca1 Rev	5'-CCACAGCTCTGCCTGAAGAT GTG-3'	Splicing: PCR amplify minigene RT sample
21	Serca Ex. 21 Fwd	5'-GTCCTCAAGATCTCACTGCCA GT-3'	Splicing: PCR amplify minigene RT sample
22	Serca Ex. 23 Rev	5'-GCCACAGCTCTGCCTGAAGAT G-3'	Splicing: PCR amplify minigene RT sample
23	Trichoplax Fwd	5'-AGGGATCCAATATTACAACTG GCAAAGATACAAGCTGG-3'	Cloning: PCR amplifying tMBNL
24	Trichoplax Rev	5'-CCGGCTCGAGCTACTGAGCTT GCTGTTGCTTTACTCG-3'	Cloning: PCR amplifying tMBNL
25	GFP_AvrII _pAC156_ R	CCTAACCGGTACGCGTCCTAGGTGCT GCTGCTTTGTAGAG	Cloning: Infusion cloning GFP into pUC256
26	GFP-Mbl_AvrII _pAC156_ F	CAA AGC AGC AGC ACC TAG GAT GGC TGC CAA C	Cloning: Infusion cloning Muclelbind into pUC256
27	Mbl_Cla1_ pAC156_R	CTAACCGGTACGCGTCCTAGATCGAT TCAAAATCTTGGCACA	Cloning: Infusion cloning Muclelbind into pUC257
28	GFP-ciMBNL_ AvrII_pAC 156_F	ACA AAG CAG CAG CAC CTA GGA TGC AGA ATC GGG CTA T	Cloning: Infusion cloning Muclelbind into pUC258
29	ciMBNL_ Cla1_pAC 156_R	GTACGCGTcctagATCGATTTCCGCGGC CGCTAT	Cloning: Infusion cloning Muclelbind into pUC259
30	GFP-ceMBNL_ AvrII_pAC 156_F	CTA CAA AGC AGC AGC ACC TAG GAT GTT TGA TGA AAA CAG TAA TGC	Cloning: Infusion cloning Muclelbind into pUC260
31	ceMBNL_ Cla1_pAC 156_R	CGGTACGCGTcctagATCGATTTAGAAC GGCGG	Cloning: Infusion cloning Muclelbind into pUC260
32	GFP-tMBNL_A vrII_pAC1 56_F	ACA AAG CAG CAG CAC CTA GGA TGA ATA TTA CAA CTG GCA AA	Cloning: Infusion cloning Muclelbind into pUC261
33	tMBNL_Cl a1_pAC15 6_R	CCGGTACGCGTCCTAGATCGATCTAC TGAGCTTG	Cloning: Infusion cloning Muclelbind into pUC262

TABLES AND FIGURES

TABLE 1: Overview of Muscleblind proteins in this study

Organism	Accession number	Protein	Similarity to Hs
Chordata *Homo sapiens*	NP_066368.2	MBNL1a	----
Chordata (Tunicata) *Ciona intestinalis*	XP_009862110.1	Zinc-Finger Protein Isoform X1 (Predicted)	47.4%
Arthropoda *Drosophila melanogaster*	NP_788390.1	Musclelbind D	62.8%
Nematoda *Caenorhabditis elegans*	NP_001257281.1	Muscleblind 1a	42.5%
Placozoa *Trichoplax adhaerens*	XP_002108472.1	Hypothetical protein (TRIADDRAFT_4444)	60.3%

Table 1 legend: Overview of Muscleblind proteins in this study. BLASTX searches using the coding region of MBNL1 isoform 41 nucleic acid sequence as query were performed to identify homologous proteins. Percent similarity of homologs and HsMBNL1 was calculated from protein alignments using EMBL-EBI EMBOSS Water alignment program, default settings used.

FIGURE 1: Zinc finger conservation and multi-species alignment

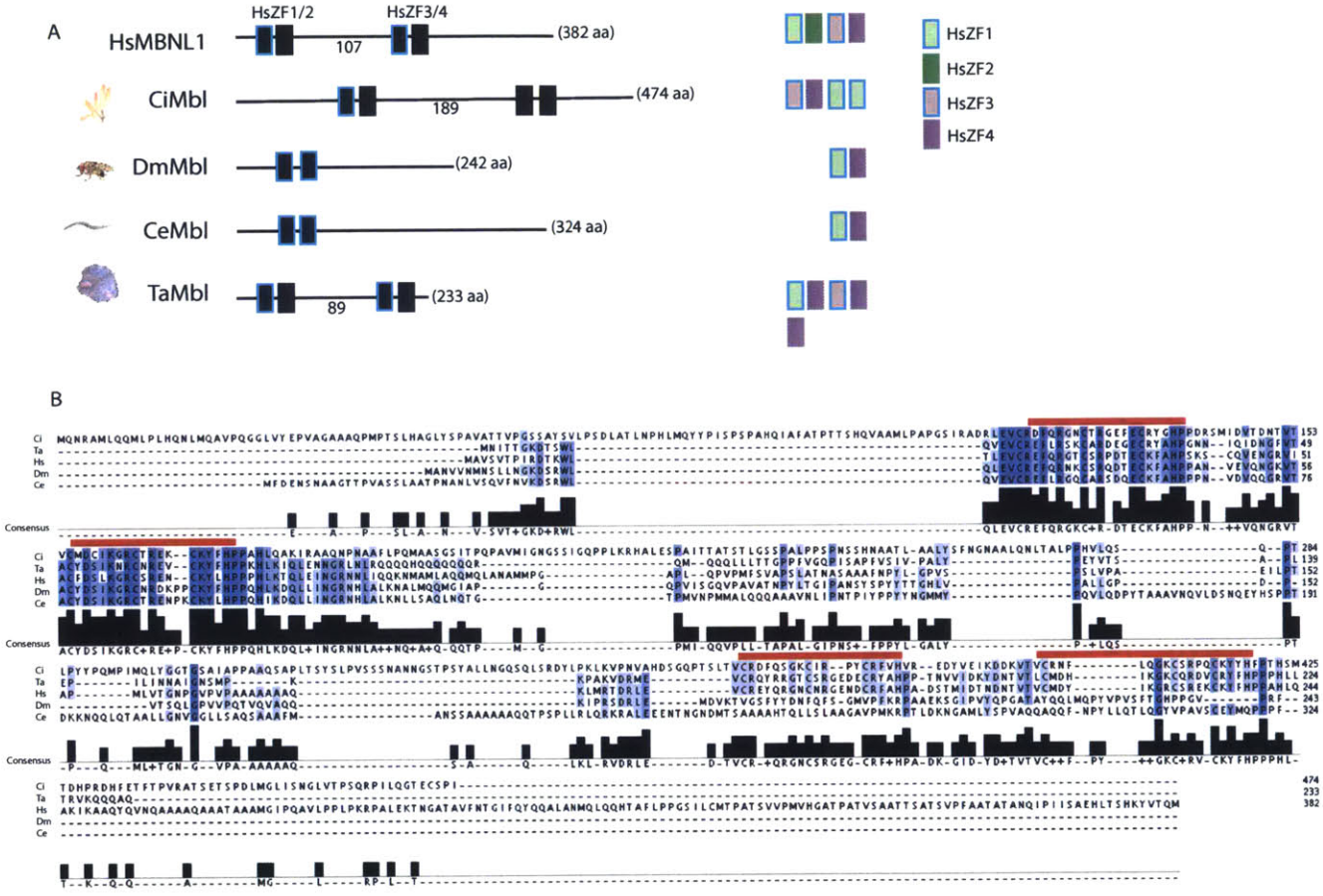


Figure 1 legend: MBNL1 homologs share the greatest extent of sequence similarity in their zinc finger domains. (**A**) Left; HsMBNL1, CiMBL, TaMBL contain two pairs of ZnFs, while the DmMBL and CeMBL proteins used contain one pair of ZnFs. Black boxes, ZnF with $CX_7CX_4CX_3H$ spacing; black boxes with cyan outline, ZnF with $CX_7CX_6CX_3H$ spacing. Spacing between tandem ZnF domains and protein length are indicated. Right; Representation of non-Hs Muscleblind ZnF similarity to HsMBNL1 ZnF1-4. Light green box, HsMBNL1 ZnF1; green, HsMBNL1 ZnF2; light purple, HsMBNL1 ZnF3; purple, HsMBNL1 ZnF4. (**B**) Multiple species alignment constructed using MUSCLE. Red bar: ZnF domains 1-4, purple shaded residues: identical residues between some or all Muscleblind homologs, black bars: height represents percentage of residues that match consensus.

FIGURE 2: Reporter assay

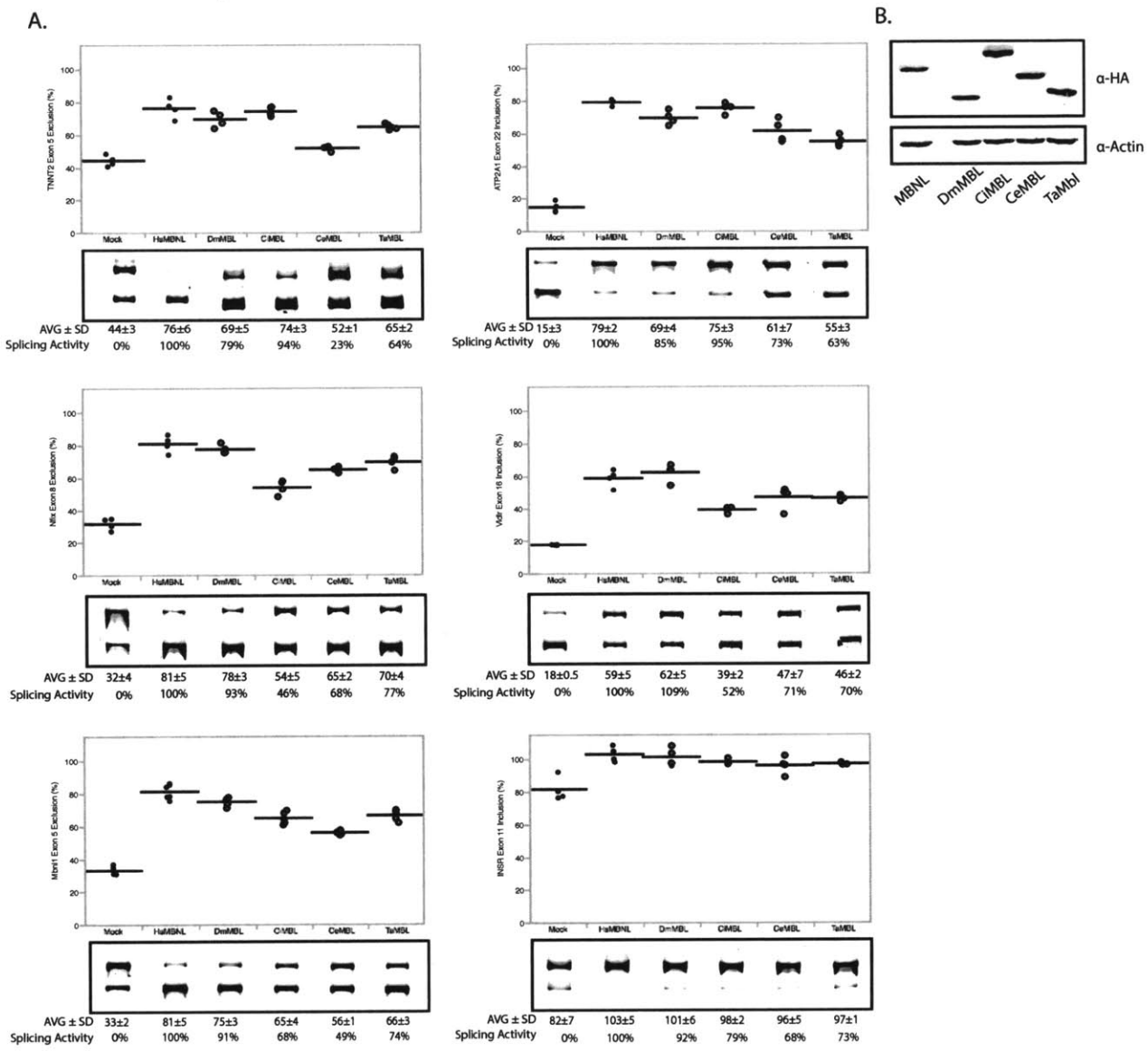
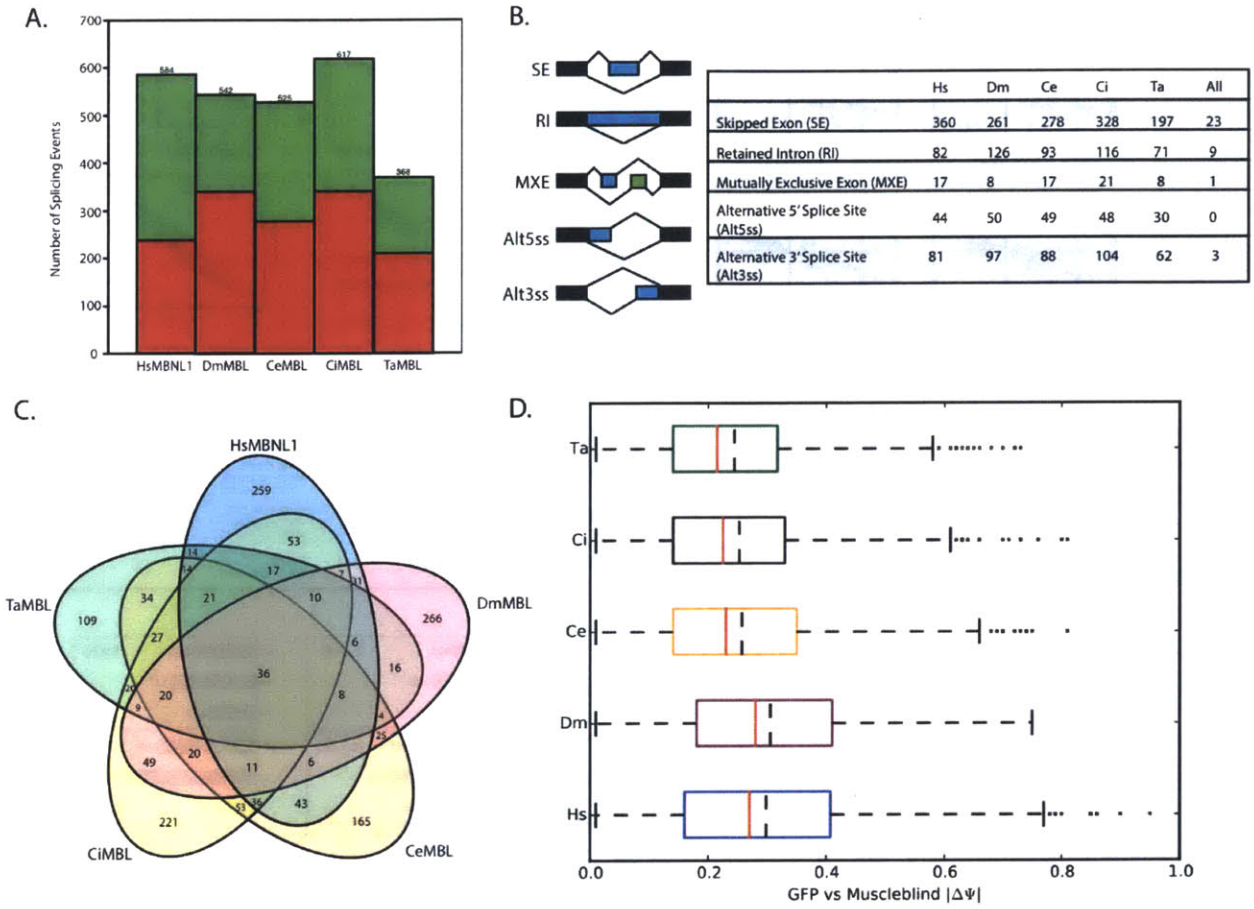


Figure 2 legend: HsMBNL1 homolog splicing regulation of reporter mini-genes in HeLa cells. **(A)** Splicing regulation of three HsMBNL1 splicing repression reporters (left) and three HsMBNL1 splicing activation reporters (right). Mock represents cells co-transfected with a splicing reporter construct and GFP-containing vector and so shows splicing of the reporter by endogenous MBNL. All other cells were co-transfected with a splicing reporter and HA-Muscleblind-expressing vector and so shows splicing of the reporter in the presence of over-expressed Muscleblind protein. Average (AVG) percent exclusion or inclusion of the alternative exon is shown with standard deviation (SD). Splicing activity relative to HsMBNL is shown. See methods for percent inclusion/exclusion and activity calculation. n ≥4. **(B)** Western blot of transiently transfected HA-tagged Muscleblind proteins confirms similar expression levels in HeLa cells. Anti-actin was used as a loading control.

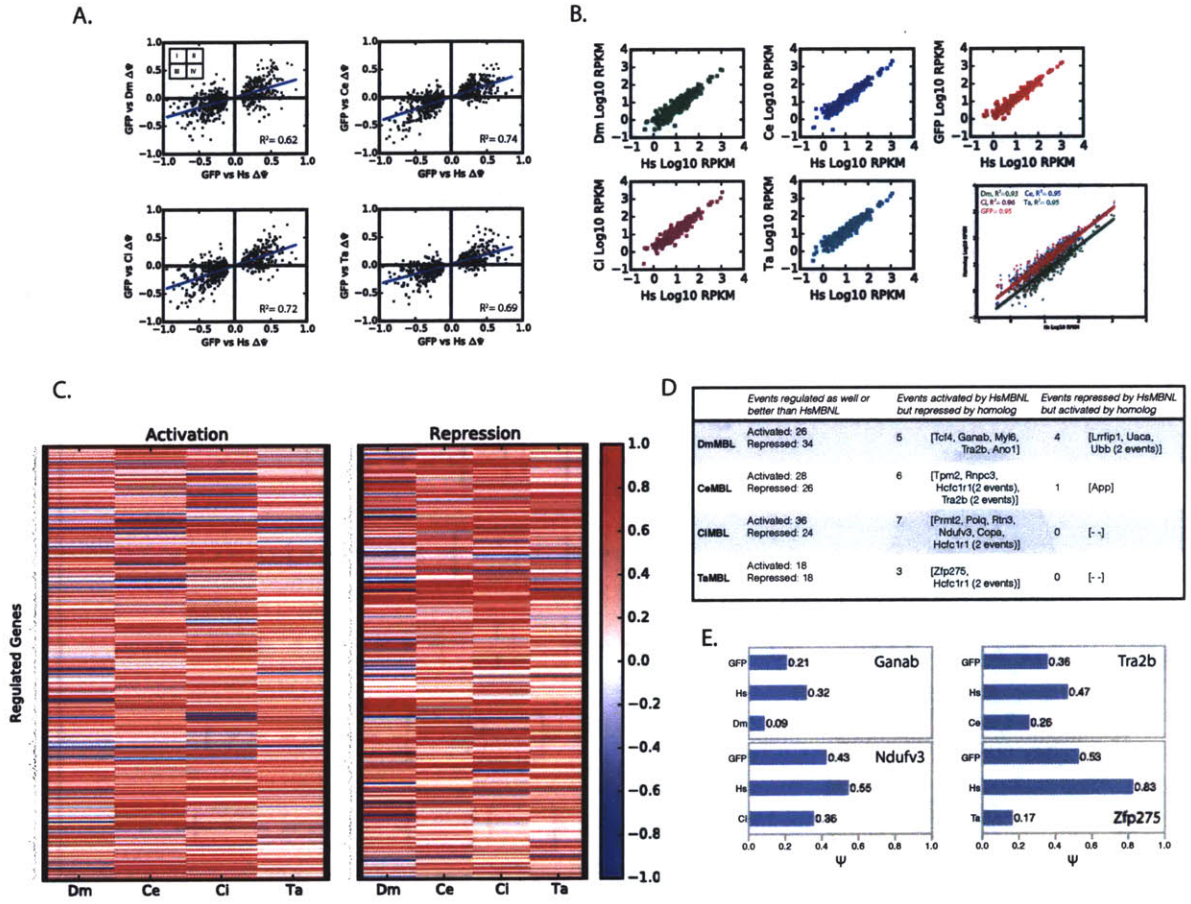
FIGURE 3: RNA-seq reveals hundreds of Muscleblind-regulated exons



	Hs	Dm	Ce	Ci	Ta	All
Skipped Exon (SE)	360	261	278	328	197	23
Retained Intron (RI)	82	126	93	116	71	9
Mutually Exclusive Exon (MXE)	17	8	17	21	8	1
Alternative 5'Splice Site (Alt5ss)	44	50	49	48	30	0
Alternative 3'Splice Site (Alt3ss)	81	97	88	104	62	3

Figure 3 legend: Hundred of alternative exons are regulated by Muscleblind homologs. (**A-C**) Spliced exons, filtered by Bayes ≥5 and $|\Delta\Psi| \geq 0.1$ (**A**) Number of exons regulated by Muscleblind proteins. Total number of exons indicated above bar. Red: repressed exons, Green: activated exons. (**B**) Significantly regulated exons broken down by alternative exon type. Hs, Dm, Ce, Ci, Ta, exons regulated in MEFs expressing: HsMBNL1, DmMBL, CeMBL, CiMBL, and TaMBL, respectively, All: events regulated in all Muscleblind cell lines. (**C**) Summary of unique or shared, regulate exons. Numbers within each section represent the maximum number of significant spliced exons, regulated by any single or combination of Muscleblind proteins. Blue ellipse: HsMBNL1-regulate exons, Pink: DmMBL-regulated exons, Orange: CeMBL-regulated exons, Yellow: CiMBL- regulated exons, Green: TaMBL-regulated exons. (**D**) Box and whisker plot showing absolute change in Ψ between GFP-expressing and Muscleblind-expressing cells for exons significantly regulated in Muscleblind-expressing cell lines ($|\Delta\Psi|$ =GFP Ψ- Muscleblind Ψ). Blue: Hs, Purple: Dm, Orange: Ce, Black: Ci, Green: Ta. Box: first to third quartiles, Red line: median, Black dashes: mean, Whiskers: Q1-1.5IQR and Q3+ 1.5IQR, Red square: out-lier.

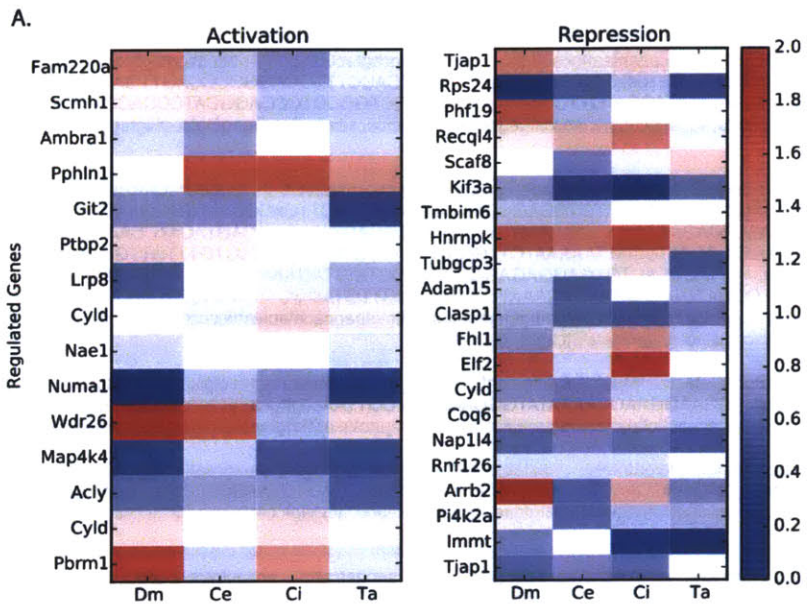
FIGURE 4: HsMBNL1 spliced exons regulated by non-mammalian homologs



	Events regulated as well or better than HsMBNL		Events activated by HsMBNL but repressed by homolog		Events repressed by HsMBNL but activated by homolog
DmMBL	Activated: 26 Repressed: 34	5	[Tcf4, Ganab, Myl6, Tra2b, Ano1]	4	[Lrrfip1, Usca, Ubb (2 events)]
CeMBL	Activated: 28 Repressed: 26	6	[Tpm2, Rnpc3, Hcfc1r1(2 events), Tra2b (2 events)]	1	[App]
CiMBL	Activated: 36 Repressed: 24	7	[Prmt2, Polq, Rtn3, Ndufv3, Copa, Hcfc1r1 (2 events)]	0	[- -]
TaMBL	Activated: 18 Repressed: 18	3	[Zfp275, Hcfc1r1 (2 events)]	0	[- -]

FIGURE 4 Legend: Many HsMBNL1-regulated exons can be regulated by non-mammalian homologs. **(A)** Pairwise ΔΨ comparison of exons regulated by HsMBNL1 and non-human Muscleblind proteins. 584 exons were significantly regulated in HsMBNL1-expressing cells compared to GFP-expressing cells (Bayes≥5 and $|\Delta\Psi| \geq 0.1$ filters). Change in splicing: ΔΨ= GFP Ψ- Muscleblind Ψ. Inset in Hs vs Dm ΔΨ graph shows quadrant numbering. **(B)** Gene expression correlation between HsMBNL1-expressing cell lines vs non-human Muscleblind- or GFP-expressing cell lines (log RPKM) of genes with exons regulated by HsMBNL1:Green: Dm, Blue: Ce, Red: GFP, Magenta: Ci, Cyan: Ta. **(C)** Fraction HsMBNL1 splicing activation and repression activity of homologs for exons significantly regulated by HsMBNL1 (Total=584, 337 activated and 247 repressed exons). Fraction activation= (non-Hs MBL Ψ- GFP Ψ)/(Hs Ψ- GFP Ψ), Fraction repression= (GFP Ψ- non-Hs MBL Ψ)/ (GFP Ψ- Hs Ψ). Color bar (right), Dark red: events regulated at least as well HsMBNL1 (fraction ≥ 1), Dark blue: oppositely regulated events (fraction < 0). **(D)** Summary of exons significantly regulated (GFP vs Hs and GFP vs Homolog Bayes≥ 5, Fraction activity ≥ 1) and oppositely regulated (GFP vs Hs and GFP vs Homolog Bayes≥ 5, Fraction activity < 0) by non-human Muscleblind, including name of genes with opposite regulation. **(E)** Psi-values for some oppositely regulated exons.

FIGURE 5: Exons regulated by all Muscleblinds



B.

Event	200nt Upstream	SE	200nt Downstream
Pbrm1	tgct gctt	cgcc (2) tcgg	gctt
Cyld	tgct (5) gctt (3) tgcc (3)	tgct (2) gctt (2) cgct	
Wdr26	tgct (4) cgcc tgcc		cgct tgcc
Numa1	tgcc gctt		tgcc tgct (3) gctt cgcc
Git 2	tgcc (2)	tgcc cgcc	tgcc (3) gctt (2) tgct cgcc (2) cgct
Pphln1	tgct (2) gctt	tgcc (2) tgct (3)	tgct (2) gctt (2) tgcc cgcc
Scmh1	tgcc gctt (2) tgct	tgcc tgct (2) gctt (2)	tgcc tgct
Fam220a	tgct		tgct gctt
Immt	tgct (4) gctt (2)		tgct
Clasp1	tgcc (3) tgct (6) cgct gctt (2)	cgcc	tgct (3) gctt tgcc
Kif3a	tgct (5) gctt (4) cgct (3)		cgct
Scaf8	tgct	tgcc tgct (2) gctt (2)	tgct (3) gctt (2) tgcc
Rps24	tgct (2) tgcc		

Figure 5 legend: Exons regulated by all Muscleblind proteins are variably including, suggesting variation in Muscleblind splicing regulation. (**A**) Fraction HsMBNL1 splicing activation and repression activity of of homologs for exons significantly regulated by all Muscleblind proteins (Total=36, 15 activated and 21 repressed exons). Fraction activation activity= (non-Hs MBL Ψ- GFP Ψ)/(Hs Ψ- GFP Ψ), Fraction repression activity= (GFP Ψ- non-Hs MBL Ψ)/ (GFP Ψ- Hs Ψ). Color bar (right), Dark red: events regulated at least as well by HsMBNL1 and a homolog (fraction ≥ 1), Dark blue: events regulated in the opposite direction by HsMBNL1 and a homolog (fraction < 0). (**B**) YGCY and GCTT 4-mers in subset of events regulated by all Muscleblind proteins. Green gene name: exon inclusion event, red gene names: exon exclusion event.

TABLE 2: YGCY/GCUU motifs in transcripts regulated by all Muscleblinds

Pbrm1	ctaattcatccaaaagtgttggatttcatatttgtaactttattttgttcattactgcaacaaaataattgcactaattgctgttatatcatgcagtactaagggtgcttattcattggtagtgcatggaggggttgttattacttagctcatgtt gcatgttaatgatgcatgtctgaaatttgttgtgtccttcaagGCATGATGGGTGGCTATCCGCCAGGCCTTCCACCTTTGCAGGGCCCAGTTGATGGCCTTGTTAGCATGGGCA GCATGCAGCCACTTCACCCTGGGGGGCCTCCACCTCACCATCTTCCGCCAGGTGTGCCTGGCCTCCCAGGCATCCCACCACCGGgtaagaacttcatcctcattc actcattaatctcatttttttttaaaaattcttttcctgtctcagccagttcttccaggaggcacgttctacccatgaggccaggctttcctcacacagaatcccagagctgtcagtagtagggactgtcctatgtgttaatcaggtgacct accaccagcagcagaggatgtactgc
Cyld	tttaaggttgtattgaatggagtctaaagttttgctttgcttgtttgccaggcactttgaattgctgtctttttacaacatggatgccaggttgctaaaatctttgctttgggacttacagcgagttcatttttttggatttttatgacagatttacta gtggctccttttgccagtcaatgttctgaaagttactgtcacaATGAGTTCAGGCCTGTGGAGCCAAGAGAAAGTTACTTCACCCTACTGGGAAGAACGGATTTTTTATCTGCTT CTTCAAGAATGCAGTGTAACAGACAAACAAACTCAGAAGCTGCTGAAAGTACCCAAAGGGAGTATAGGACAGTACATCCAAGACCGTTCTGTGGGGCATT CAAGAGTTCCTTCCACAAAAGGCAAGAAAAATCAGATTGGATTAAAAATCTTGGAGCAACCGCATGCAGTTCTGTTTGTTGATGAAAAGGATGTTGTAGAAA TAAATGAAAAATTCACAGAGTTACTGTTGGCAATTACCAACTGTGAGGAGAGGCTCAGCCTATTTAGAAACAGACTCCGACTAAGTAAAGGCCTCCAGGTA GACGTGGGCAGTCCTGTGAAAGTACAGCTGCGATCTGGGGAAGAGAAATTTCCAGGAGTTGTACGCTTCAGAGGACCTTTATTAGCGGAGAGGACGGTG TCGGGGATTTTCTTTGGAGTAGAATTATTGgtaagtttgaaaagacattttagtgtttttgtgagtgtgtgtgttgtgtgagcacatacatatttccccctttgtatgcagtatatttttttatactcatgatccttaggatc aatggtttttgtttgttttggtgtacttctcaggatagaccccaggaccctctaccactgatctatatctcagccccaaggttt
Wdr26	atcatcattttgggattatcctaatacctagcactgtatatagaatgtgctatactaacatgaactgtgtctgatggtgagggacggaagaaactaaaacgccctgtgctgcctttgtgccgagtagtgtgctcagagtatttgctc tcacatatgttcatggtcttgtaagagaaacatttttatttttcatttttatagGGAAGGAACCAAAATATGGCAGAGCTCGTTAActtgttgaacgctgccatttcggcaaatctaggtctggaaagttttccccacca ccttgatacttgaagatgaagaaagtcataaagctaacagatgaggacaaattcttgattaaatataatagtactagagagaagtaataaatttatgtaaatatctttctttttcctcaaagagaacaagtaaaaacgtctcgt
Numa1	tatatatgtaaagatggccttgcactcttgatcttcctgcctttcctcccaagtgttgtgatctcaggcttaatatggcctgttacctgaatccagttggcctacactcccattctctgggtgaagtatcagagcccaagtgaggtcagtc tggcagtggctgcatctcactgtctccattacttctccccatgcacagGTGGAGCAACTAGAGGTATTTCAAAGAGAGCAGACTAAGCAGgtaatgcctggggttctcgataaatgctggctgcttaa gcggtgaccagtctcggtgcatgaccccggctgcagcttgctccgccagctcctgttgtgaactgtttgtgaaaggcagggtcaagccatggagtccaggcccaggctacccaggagtacagtgttacaggccctctctgtgg cctcatcctacccatccca
Git2	agctacgatggcggtaggatctcttccttaagcacaagtggcctcttcagtgcagagtggtttccagcaggccactgcctaattagggatcattgccacagctgtcaccgagctgagggacatgtgtgagctgtgcagtggga ggtggggacgtgtttgttttatgttttcccaagcgtgtttctctctccacttccagATTGGGAGGAGTGCACTTGTGACCTCCTCTTCGTCTCTGCCCTCCTTCCCCTCCACACTTTCCTG GTCCAGGGACGAAAGCGCCCGAAGGgttaagtacactctgactggtctgccctggaggagggccagcacccctggcttgtgccccgccccttcctaactgcccctctgtcctgcatgaagcagctttgc tcctgggttctcatagcagcaggcacatgcagccgtgcacatccggcctggagagagctggccttcctctgggcctccacgctgtgcgccttcc
Pphln1	ctcattgaaatattttatttccattaatctatatgaagatgttgctattagctatagccacatttttcattcgtgtgtttttgttggtttgcatttactctgttgaattatggttaagaactcctgtgcttataagaaagttttccaaaaccaaataga gaaattagtgatgtctctatatttcttttgtctgttttagGATGGCTACCATAGAGTAGTTAATGTTGTGCCAAAGAGACCACCTCTGCTAGACAAGAGACCACCTCTGCTAGACA AGAGACCACCTCTGCTAGCCAGACCTGATGAAGGAGGCTACAGTAGATATTACAGTCATGTTGATTGCCGAGTATGTGACGAGGGCCGCAGTTTTTCTCA TGATCGAAGAAGTGGCCCATCTCACAGTGGAgtatgtaaatttcccccgtgtactaattgttctttctttcaagtttaaaatccatgtaactaaagtctatgtgcaggtctctgaagtggagtcatggtgcttgtgtta gggtgaaaggagagcctgatgcttgtgttctgcctccctccccccaccccaactccctttttacactagggtctcactgtatagtcccgccct
Scmh1	gaaaacaaaacacaagacctgtgccttgatttctgttctgtgttgttctcctaatggcttttgtttgttctgagttcttggtggtgcagcttggagccaagtggagccaaggccagagttgagcagatggaagagagctgctctctg acttctctctaccaggcacactctaaggttcctgggttctttgtttcctgtagTTTGTATCTACTTGAACAAGAGCGGCAGCACGGGCCCCCACCTGGATAAGAAGAAGATCCAACAAC TCCCTGACCATTTTGGGCCAGCCCGTGCCTCTGTGGTGCTGCAGCAGGCTGTCCAGGCTTGCATTGACTGTGCTTATCACCAGAAAACTGTCTTCAGCT TCCTCAAACAGGGCCACGGCGGTGAAGTCATTTCAGgtaaacttttgggctagtttggcctcctaacccgagtgtcagaccttcattctgatgagagatcctagtactgacttcatactcataggctctgga gaatgcaaacgtatctcattcctccagagtccatactgccagttcccaggctctgctggacttatataacctgaatacccatatagtatgtaccaaatacctctg
Fam220a	gtgttcaattctctgtgtaccctctcgagtctaggctattctaattgctagcctgaaattttgtgtattctgttttacactagtcaattagaggcatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtaaaatgtgtatgatata tattatgtgtatgtatggatctatgtgtgattatctgcagGTGGTGGTGGAAGAGGAATAGgaggaagaggaggaggaggaggaggaagaggcggaggcggcagccagaagattctggtcactggatcctctg gagctggagttgcaggcagttgtgaactgcttgagtccaggtctctgcaagaacagcaagcacccttaacagctgagccatctctctggctccctagagaagtcttctgaacatatctgtcatgtctt
Immt	cagagaaaccctgtctcgaaaaaccaaaataaaaaaagaaaaagaaaaagaactgctgtctctgactctgtaaatataggaactaaacatttgctttctagaaaccctctttcttggtctttctggatgttgctatttaatgatgg ctaaccaaagtgcgtgcttttccttttcacactattcttctctttctttgtagGTATTTCTGACCTAGgtgagtaattatgtatattctatgtttaatggcatttttgggggggtgttgaaatagtgaactctttctggtgtgtgtgcatgtgt gcatgggtgcgcgtgtgctacccatctaaactatagcaagtagaacctgaatgtgtcataattctcaattgtcatccatagataatcaagaataattgtgttaagtgtaaggct
Clasp1	tgggacagagcctgaaggactaggacatgccattcagaccgcagcgtgctacgatctactccaaaccgcttgctgacctgtgtttttggttgctaagagatgttgctatacactttttgttgccagaacaatccccacatatgct tctgacataattctctatgctttttccccatccttcaatattttcctgcctacagCTGTTAGGAGACGCCAGGAGCAAGgtacccgtgctttaacttttgtttacaactttttttttctgttgcaagtctatttctttttagaatttag ttgaacgcagcacaaaatgtagttgtagttagagactctggtctcagtgtttctggtaagcattctgatttgctgccgtgtgatggagtcccggcccctctcatccttctgggaaggtgctctgttttta
Kif3a	ccgaacgaacatgcttctttttttctgtcggctctgtttcagttttgattatttgattcatgtctgtgctcgtgctaaatgagtgctgttacacgttttctctgttttacaaaggaaataatgttttcctctcttccctccccatctctgtgcacacg cttcctgtctgcggctcgcttaccttacgctgctttgcaagACCAAGCAGgttggtgtgattggctggcacttcacctggtgcgctcactcattttctcaagaccttttatttaagtacttgacttttttatttttaataatttttacaagg taagtatctgtctttttaaaacaattccattttagctatttttaaaagccttctcccctaaatccttaaagacctattaagtcagttttgaaaatgtac
Scaf8	actttattactttcatatgtgatttcagatatctctttctagggccagatttgctgttaatggacagttcctaactcatagtagttatgacaaaaatgtgatcaaccatgaccttgaacactggctgagctcagactctagtgagctcatttt gtctgttttgttttctcaacatctttctctactcctctctctcagGATTGTGCCTCTACTCCTGCTTGTACAGACTGCTTTCCAGCAGCATGGACACAAGAAGCATAgtaagatgactatacccc acataaggcaggctctggggaaatgttaggtctaatgtggtagtgaaattctgtattgcttgtatgggtagtacttgggaagcacatagagttgccacttagaacctattagtacttttataagaatcatacagggaataaaactta atgcttttataagaaatgtgctatttctttaaa
Rps24	actttgttaacattggaactggtaatcagggaaggtgctgtcagaagtgggtgctgtgcaaggaatgttctgtacacaacagaggacagtcacctgaggtatctgcatgtacttcagataattttggtgtttgaaaaaggactact aatgaactgtctgttttcctccttcccattccccctttcctgccactgattcagTGAgtggagattggatacaggtatagttcaagctattccgtagatcactagactccttgttactgatgtagtagaccaagcacgtgaacatttataa gtagattggcaagtgtagtttataacagtactaacattgactgtctaaacttcagagatggtgacttgttttgtttcatttctaacgtgaatgttagtttagttt

Table 2 legend: Putative HsMBNL1 binding sites in events regulated by all Muscleblind homologs. Left Column, Green: Gene names of inclusion isoforms, Red: Gene names of exclusion isoforms. Right column, lowercase: intronic sequence, uppercase: exons, bold and orange: CGCC, TGCT/GCTT, CGCT, TGCC putative MBNL-binding motifs, underlined: GC di-nucleotides within putative MBNL-binding motifs, purple: 3' splice site, blue: 5' splice site.

FIGURE 6: Subcellular localization of Muscleblind proteins

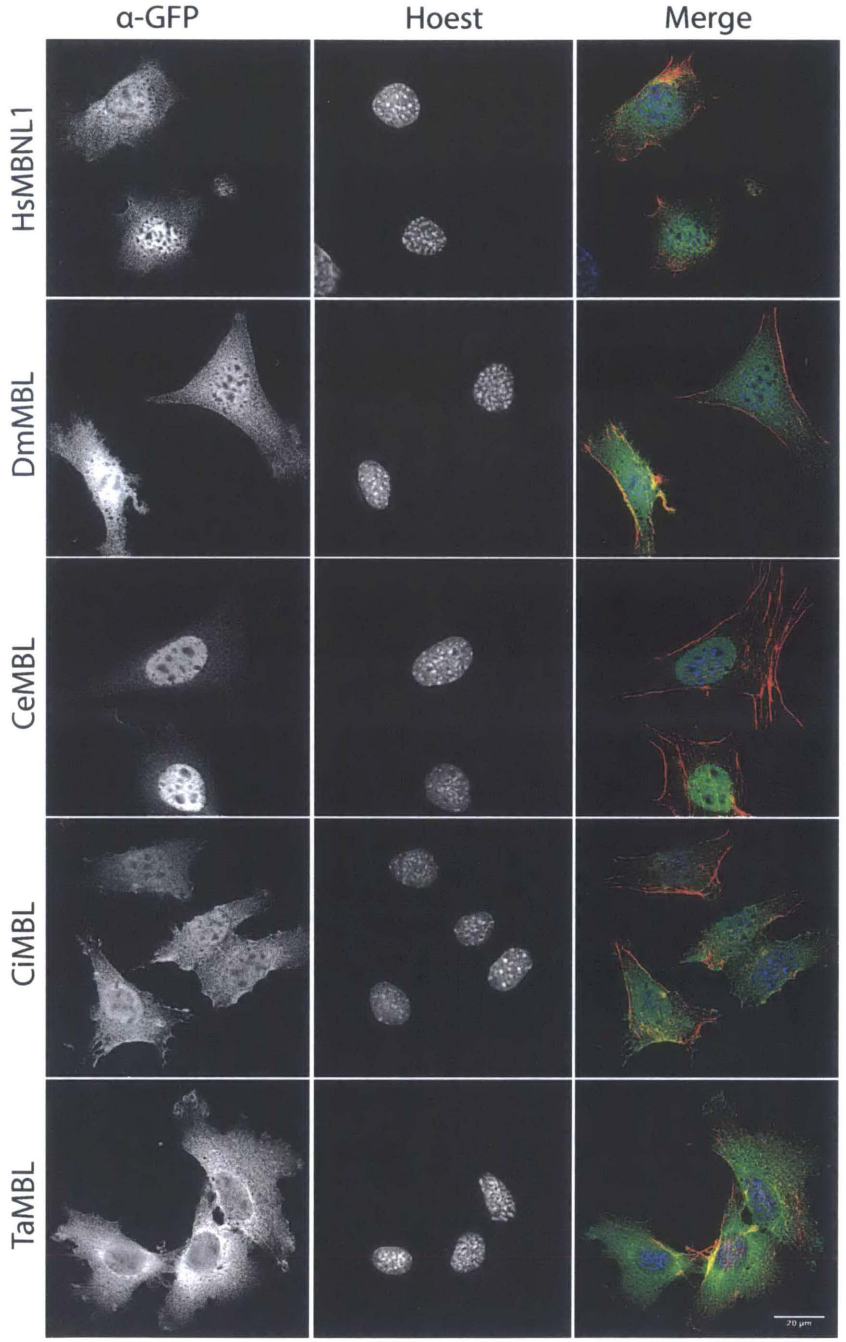


Figure 6: Subcellular localization of Muscleblind proteins. MEF cells stably expressing GFP-tagged Muscleblind proteins were fixed after 24 hours in culture and imaged with a fluorescence microscope. Rows represent cells expressing the different Muscleblind proteins. anti-GFP; gray-scale images of Muscleblind proteins, Hoest; gray-scale images of the nucleus, Merge; Muscleblind proteins (green), nucleus(blue), and actin(red). All images were taken with a 60X objective, scale bar; bottom right, 20uM.

CITATIONS

1. McManus, C. J., & Graveley, B. R. (2011). RNA structure and the mechanisms of alternative splicing. Curr Opin Genet Dev, 21(4), 373-379. doi: 10.1016/j.gde.2011.04.001

2. Lee, Y., & Rio, D. C. (2015). Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annu Rev Biochem. doi: 10.1146/annurev-biochem-060614-034316

3. Pascual, M., Vicente, M., Monferrer, L., & Artero, R. (2006). The Muscleblind family of proteins: an emerging class of regulators of developmentally programmed alternative splicing. Differentiation, 74(2-3), 65-80. doi: 10.1111/j.1432-0436.2006.00060.x

4. Fardaei et al. (2002). Three proteins, MBNL, MBLL, and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum Mol Genetics* 11, 7: 805-814

5. Konieczny, P., Stepniak-Konieczna, E., & Sobczak, K. (2014). MBNL proteins and their target RNAs, interaction and splicing regulation. Nucleic Acids Res, 42(17), 10873-10887. doi: 10.1093/nar/gku767

6. Poulos, M. G., Batra, R., Li, M., Yuan, Y., Zhang, C., Darnell, R. B., & Swanson, M. S. (2013). Progressive impairment of muscle regeneration in muscleblind-like 3 isoform knockout mice. Hum Mol Genet, 22(17), 3547-3558. doi: 10.1093/hmg/ddt209

7. Squillace, R. M., Chenault, D. M., & Wang, E. H. (2002). Inhibition of Muscle Differentiation by the Novel Muscleblind-Related Protein CHCR. Developmental Biology, 250(1), 218-230. doi: 10.1006/dbio.2002.0798

8. Lee, K. S., Cao, Y., Witwicka, H. E., Tom, S., Tapscott, S. J., & Wang, E. H. (2010). RNA-binding protein Muscleblind-like 3 (MBNL3) disrupts myocyte enhancer factor 2 (Mef2) {beta}-exon splicing. J Biol Chem, 285(44), 33779-33787. doi: 10.1074/jbc.M110.124255

9. Kino, Y., Washizu, C., Kurosawa, M., Oma, Y., Hattori, N., Ishiura, S., & Nukina, N. (2015). Nuclear localization of MBNL1: splicing-mediated autoregulation and repression of repeat-derived aberrant proteins. Hum Mol Genet, 24(3), 740-756. doi: 10.1093/hmg/ddu492

10. Tran, H., Gourrier, N., Lemercier-Neuillet, C., Dhaenens, C. M., Vautrin, A., Fernandez-Gomez, F. J., . . . Sergeant, N. (2011). Analysis of exonic regions involved in nuclear localization, splicing activity, and dimerization of Muscleblind-like-1 isoforms. J Biol Chem, 286(18), 16435-16446. doi: 10.1074/jbc.M110.194928

11. Osborne, R. J., Lin, X., Welle, S., Sobczak, K., O'Rourke, J. R., Swanson, M. S., & Thornton, C. A. (2009). Transcriptional and post-transcriptional impact of toxic RNA in myotonic dystrophy. Hum Mol Genet, 18(8), 1471-1481. doi: 10.1093/hmg/ddp058

12. Masuda, A., Andersen, H. S., Doktor, T. K., Okamoto, T., Ito, M., Andresen, B. S., & Ohno, K. (2012). CUGBP1 and MBNL1 preferentially bind to 3' UTRs and facilitate mRNA decay. Sci Rep, 2, 209. doi: 10.1038/srep00209

13. Wang, E. T., Cody, N. A., Jog, S., Biancolella, M., Wang, T. T., Treacy, D. J., . . . Burge, C. B. (2012). Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. Cell, 150(4), 710-724. doi: 10.1016/j.cell.2012.06.041

14. Du, H., Cline, M. S., Osborne, R. J., Tuttle, D. L., Clark, T. A., Donohue, J. P., . . . Ares, M., Jr. (2010). Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. Nat Struct Mol Biol, 17(2), 187-193. doi: 10.1038/nsmb.1720

15. Adereth, Y., Dammai, V., Kose, N., Li, R., & Hsu, T. (2005). RNA-dependent integrin alpha3 protein localization regulated by the Muscleblind-like protein MLP1. Nat Cell Biol, 7(12), 1240-1247. doi: 10.1038/ncb1335

16. Rau, F., Freyermuth, F., Fugier, C., Villemin, J.-P., Fischer, M.-C., Jost, B., . . . Charlet-Berguerand, N. (2011). Misregulation of miR-1 processing is associated with heart defects in myotonic dystrophy. Nat Struct Mol Biol, 18(7), 840-845. doi: http://www.nature.com/nsmb/journal/v18/n7/abs/nsmb.2067.html#supplementary-information

17. Kanadia, R. N., Urbinati, C. R., Crusselle, V. J., Luo, D., Lee, Y.-J., Harrison, J. K., . . . Swanson, M. S. (2003). Developmental expression of mouse muscleblind genes Mbnl1, Mbnl2 and Mbnl3. Gene Expression Patterns, 3(4), 459-462. doi: http://dx.doi.org/10.1016/S1567-133X(03)00064-4

18. Begemann, G., Paricio, N., Artero, R., Kiss, I., Perez-Alonso, M. and Mlodzik, M. (1997) muscleblind, a gene required for photoreceptor differentiation in Drosophila, encodes novel nuclear Cys3His-type zinc-finger-containing proteins. Development 124, 4321–4331.

19. Ho, T. H., Charlet-B, N., Poulos, M. G., Singh, G., Swanson, M. S., & Cooper, T. A. (2004). Muscleblind proteins regulate alternative splicing. EMBO J 23, 3103-3112.

20. Goers, E. S., Voelker, R. B., Gates, D. P., & Berglund, J. A. (2008). RNA binding specificity of Drosophila muscleblind. Biochemistry, 47(27), 7284-7294. doi: 10.1021/bi702252d

21. Lambert, N., Robertson, A., Jangi, M., McGeary, S., Sharp, P. A., & Burge, C. B. (2014). RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. Mol Cell, 54(5), 887-900. doi: 10.1016/j.molcel.2014.04.016

22. Cass, D., Hotchko, R., Barber, P., Jones, K., Gates, D. P., & Berglund, J. A. (2011). The four Zn fingers of MBNL1 provide a flexible platform for recognition of its RNA binding elements. BMC Mol Biol, 12, 20. doi: 10.1186/1471-2199-12-20

23. Teplova, M., & Patel, D. J. (2008). Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1. Nat Struct Mol Biol, 15(12), 1343-1351. doi: 10.1038/nsmb.1519

24. Mooers, B. H., Logue, J. S., & Berglund, J. A. (2005). The structural basis of myotonic dystrophy from the crystal structure of CUG repeats. Proc Natl Acad Sci U S A, 102(46), 16626-16631. doi: 10.1073/pnas.0505873102

25. Warf, M. B., & Berglund, J. A. (2007). MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. RNA, 13(12), 2238-2251. doi: 10.1261/rna.610607

26. Merkin et al. (2012). Evolutionary Dynamics of Gene and Isoform Regulation in Mammalian Tissues. Science 21, 338: 1593-1599.

27. Fernandez-Costa, J. M., & Artero, R. (2010). A conserved motif controls nuclear localization of Drosophila Muscleblind. Mol Cells, 30(1), 65-70. doi: 10.1007/s10059-010-0089-9

28. Irion U. 2012. Drosophila muscleblind Codes for Proteins with One and Two Tandem Zinc Finger Motifs. PLoS ONE 7(3): e34248. doi:10.1371/journal.pone.0034248

29. Vicente-Crespo, M., Pascual, M., Fernandez-Costa, J. M., Garcia-Lopez, A., Monferrer, L., Miranda, M. E., . . . Artero, R. D. (2008). Drosophila muscleblind is involved in troponin T alternative splicing and apoptosis. PLoS One, 3(2), e1613. doi: 10.1371/journal.pone.0001613

30. Artero, R. et al. (1998). The muscleblind Gene Participates in the Organization of Z-Bands and Epidermal Attachmentsof Drosophila Muscles and is Regulated by Dmef2. Dev Biol. 15, 195(2): 131-143.

31. Fernandez-Costa, J. M., Llamusi, M. B., Garcia-Lopez, A., & Artero, R. (2011). Alternative splicing regulation by Muscleblind proteins: from development to disease. Biol Rev Camb Philos Soc, 86(4), 947-958. doi: 10.1111/j.1469-185X.2011.00180.x

32. de Haro, M., Al-Ramahi, I., De Gouyon, B., Ukani, L., Rosa, A., Faustino, N. A., . . . Botas, J. (2006). MBNL1 and CUGBP1 modify expanded CUG-induced toxicity in a Drosophila model of myotonic dystrophy type 1. Hum Mol Genet, 15(13), 2138-2145. doi: 10.1093/hmg/ddl137

33. Picchio, L., Plantie, E., Renaud, Y., Poovthumkadavil, P., & Jagla, K. (2013). Novel Drosophila model of myotonic dystrophy type 1: phenotypic characterization and genome-wide view of altered gene expression. Hum Mol Genet, 22(14), 2795-2810. doi: 10.1093/hmg/ddt127

34. Yu, Z., Teng, X., & Bonini, N. M. (2011). Triplet repeat-derived siRNAs enhance RNA-mediated toxicity in a Drosophila model for myotonic dystrophy. PLoS Genet, 7(3), e1001340. doi: 10.1371/journal.pgen.1001340

35. Monferrer, L., & Artero, R. (2006). An interspecific functional complementation test in Drosophila for introductory genetics laboratory courses. J Hered, 97(1), 67-73. doi: 10.1093/jhered/esj003

36. Sasagawa, N., Ohno, E., Kino, Y., Watanabe, Y., & Ishiura, S. (2009). Identification of Caenorhabditis elegans K02H8.1 (CeMBL), a functional ortholog of mammalian MBNL proteins. J Neurosci Res, 87(5), 1090-1097. doi: 10.1002/jnr.21942

37. Wang, L. C., Hung, W. T., Pan, H., Chen, K. Y., Wu, Y. C., Liu, Y. F., & Hsiao, K. M. (2008). Growth-dependent effect of muscleblind knockdown on Caenorhabditis elegans. *Biochem Biophys Res Commun, 366*(3), 705-709. doi: 10.1016/j.bbrc.2007.12.024

38. Spilker, K. A., Wang, G. J., Tugizova, M. S., & Shen, K. (2012). Caenorhabditis elegans Muscleblind homolog mbl-1 functions in neurons to regulate synapse formation. Neural Dev, 7, 7. doi: 10.1186/1749-8104-7-7

39. Wang, L. C., Chen, K. Y., Pan, H., Wu, C. C., Chen, P. H., Liao, Y. T., . . . Hsiao, K. M. (2011). Muscleblind participates in RNA toxicity of expanded CAG and CUG repeats in Caenorhabditis elegans. Cell Mol Life Sci, 68(7), 1255-1267. doi: 10.1007/s00018-010-0522-4

40. Passamaneck, Y. J., & Di Gregorio, A. (2005). Ciona intestinalis: chordate development made simple. Dev Dyn, 233(1), 1-19. doi: 10.1002/dvdy.20300

41. Srivastava, M., Begovic, E., Chapman, J., Putnam, N. H., Hellsten, U., Kawashima, T., . . . Rokhsar, D. S. (2008). The Trichoplax genome and the nature of placozoans. Nature, 454(7207), 955-960. doi: 10.1038/nature07191

42. Goers, E. S., Purcell, J., Voelker, R. B., Gates, D. P., & Berglund, J. A. (2010). MBNL1 binds GC motifs embedded in pyrimidines to regulate alternative splicing. Nucleic Acids Res, 38(7), 2467-2484. doi: 10.1093/nar/gkp1209

43. Miller, J.W., Urbinati, C.R., Teng-Umnuay, P., Stenberg, M.G., Byrne, B.J., Thornton, C.A., and Swanson, M.S. (2000). Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. EMBO J. 19, 4439–4448

44. Botta, A., Malena, A., Tibaldi, E., Rocchi, L., Loro, E., Pena, E., . . . Vergani, L. (2013). MBNL142 and MBNL143 gene isoforms, overexpressed in DM1-patient muscle, encode for nuclear proteins interacting with Src family kinases. Cell Death Dis, 4, e770. doi: 10.1038/cddis.2013.291

45. Grammatikakis, I., Goo, Y. H., Echeverria, G. V., & Cooper, T. A. (2011). Identification of MBNL1 and MBNL3 domains required for splicing activation and repression. Nucleic Acids Res, 39(7), 2769-2780. doi: 10.1093/nar/gkq1155

46. Purcell, J., Oddo, J. C., Wang, E. T., & Berglund, J. A. (2012). Combinatorial mutagenesis of MBNL1 zinc fingers elucidates distinct classes of regulatory events. Mol Cell Biol, 32(20), 4155-4167. doi: 10.1128/MCB.00274-12

47. Gates, D. P., Coonrod, L. A., & Berglund, J. A. (2011). Autoregulated splicing of muscleblind-like 1 (MBNL1) Pre-mRNA. J Biol Chem, 286(39), 34224-34233. doi: 10.1074/jbc.M111.236547

48. Hino, S., Kondo, S., Sekiya, H., Saito, A., Kanemoto, S., Murakami, T., . . . Imaizumi, K. (2007). Molecular mechanisms responsible for aberrant splicing of SERCA1 in myotonic dystrophy type 1. Hum Mol Genet, 16(23), 2834-2843. doi: 10.1093/hmg/ddm239

49. Sen, S., Talukdar, I., Liu, Y., Tam, J., Reddy, S., & Webster, N. J. (2010). Muscleblind-like 1 (Mbnl1) promotes insulin receptor exon 11 inclusion via binding to a downstream evolutionarily conserved intronic enhancer. J Biol Chem, 285(33), 25426-25437. doi: 10.1074/jbc.M109.095224

50. Vicente, M., Monferrer, L., Poulos, M. G., Houseley, J., Monckton, D. G., O'Dell K, M., . . . Artero, R. D. (2007). Muscleblind isoforms are functionally distinct and regulate alpha-actinin splicing. Differentiation, 75(5), 427-440. doi: 10.1111/j.1432-0436.2006.00156.x

51. Katz, Y., Wang, E. T., Airoldi, E. M., & Burge, C. B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Meth, 7*(12), 1009-1015. doi: http://www.nature.com/nmeth/journal/v7/n12/abs/nmeth.1528.html#supplementary-information

52. Tacke et al. 1998. Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell* 93, 1:139-148.

53. Matera, A. G., & Wang, Z. (2014). A day in the life of the spliceosome. Nat Rev Mol Cell Biol, 15(2), 108-121. doi: 10.1038/nrm3742

54. Terenzi, F., & Ladd, A. N. (2014). Conserved developmental alternative splicing of muscleblind-like (MBNL) transcripts regulates MBNL localization and activity. RNA Biology, 7(1), 43-55. doi: 10.4161/rna.7.1.10401

55. Philips, A.V., Timchenko, L.T. and Cooper, T.A. (1998) Disruption of Splicing Regulated by a CUG-Binding Protein in Myotonic Dystrophy. Science, 280, 737- 741.

56. Ladd, A. N., Charlet, N., & Cooper, T. A. (2001). The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. Mol Cell Biol, 21(4), 1285-1296. doi: 10.1128/MCB.21.4.1285-1296.2001

57. Kosaki, A., Nelson, J. and Webster, N.J.G. (1998) Identification of Intron and Exon Sequences Involved in Alternative Splicing of Insulin Receptor Pre-mRNA. J. Biol. Chem., 273, 10331-10337.

58. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29(1), 15-21. doi: 10.1093/bioinformatics/bts635

59. Machuca-Tzili, L. et al. (2006). Flies deficient in Muscleblind protein model features of myotonic dystrophy with altered splice forms of Z-band associated transcripts. Hum. Genetics 120, 4:487-499.

60. Panaite, PA. et al. (2008). Myotonic dystrophy transgenic mice exhibit pathologic abnormalities in diaphragm neuromuscular junctions and phrenic nerves. Journal of neuropathology and experimental neurology 67:763–772.

61. Berna, L., & Alvarez-Valin, F. (2014). Evolutionary genomics of fast evolving tunicates. Genome Biol Evol, 6(7), 1724-1738. doi: 10.1093/gbe/evu122

62. Hughes, AL. and Friedman, R. (2005). Loss of ancestral genes in the genomic evolution of Ciona Intestinalis. Evol. Dev. 7, 3:196-200.

63. Boutz, P. L., Stoilov, P., Li, Q., Lin, C. H., Chawla, G., Ostrow, K., . . . Black, D. L. (2007). A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. Genes Dev, 21(13), 1636-1652. doi: 10.1101/gad.1558107