EXPLORING THE MUSCLEBLIND SEQUESTRATION MODEL FOR RNA-MEDIATED DISEASES

By

YUAN YUAN

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2007

To the loving and the beloved

献给我的父母

ACKNOWLEDGMENTS

I would like to thank my mentor, Dr. Maurice Swanson, for giving me this exciting project to work on and superb intellectual support and guidance during the past six years. To think creatively and to attack scientific problems meticulously, are the two qualities he greatly encouraged and reinforced in me, which will benefit me for the rest of my scientific career. I would also like to thank the past and present members of the Swanson lab, in particular Dr. Keith R. Nykamp, Dr. Patana Teng-Umnuay, Myrna Stenberg, Mike Poulos, Jason O'Rourke and Jihae Shin, for the fun intellectural discussions as well as their technical help. I would like to extend a special thanks to Dr. Jack Griffith and Dr. Sarah Compton for hosting me at University of North Carolina at Chapel Hill and teaching me electron microscopy. I also greatly appreciate the help of Dr. Krzysztof Sobczak for performing the RNA structural probing described in this dissertation. I would also like to thank my committee members, Dr. Alfred Lewin, Dr. John Aris, Dr. Richard Condit and Dr. Brian Burke, for providing me with valuable assistance during my graduate career. I am sincerely grateful to them. Last, but not least, I would like to thank my parents for their unconditional love and support, for which I will be forever indebted.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CDM	Congenital myotonic dystrophy
CLCN1	Skeletal muscle chloride channel 1
cTNT/TNNT2	Cardiac troponin T
CUGBP1	CUG repeats RNA binding protein 1
DM	Myotonic dystrophy
DM1	Myotonic dystrophy type I
DM2	Myotonic dystrophy type II
DMPK	Dystrophia myotonica protein kinase
DMWD	Dyotrophia myotonica-containing WD repeat motif
dpc	Days post coitus
EM	Electron microscopy
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
F exon	Fetal exon
FTDP-17	Frontotemporal dementia with parkinsonism linked to chromosome 17
HD	Huntington disease
Hfq	Host factor 1; E.coli homolog to the Sm protein
hnRNP	Heterogeneous ribonucleoproteins
HSA	Human skeletal muscle actin
HSA^{LR}	Transgenic mice carrying 250 CTG repeats in the 3-UTR of HSA transgene

HSA^{SR}	Transgenic mice carry 5 CTG repeats in the 3-UTR of HSA transgene
IR	Insulin receptor
ISE	Intronic splicing enhancer
ISS	Intronic splicing silener
MAPT	Microtubule-associated protein tau
Mbl	Muscleblind protein
MBNL	Muscleblind-like protein
Nab2p	Nuclear polyadenylated-RNA binding protein 2
Nova	Neuro-oncological ventral antigen
nPTB	Neuronal polypyrimidine tract binding protein
NMD	Nonsense-mediated decay
PTB	Polypyrimidine tract binding protein
PTC	Premature termination codon
RARγ	Retinoic acid receptor gamma
SCA	Spinocerebellar ataxias
Six5	Sine oculis-related homeobox 5 homolog
snRNP	Small nuclear ribonucleoprotein
T5	The 500 nt region including Tnnt3 fetal exon and the upstream intronic sequence
T5.1	The most 5’ 125 nt region of T5; it is entirely intronic
T5.45	The most 3’ 200 nt region of T5; it includes the entire Tnnt3 fetal exon and 161 nt of the upstream intronic sequence

Tnnt3	Fast troponin T in skeletal muscle
TTP	Tristetraprolin
U2AF	U2 auxiliary factor
UTR	Untranslated region
ZNF9	Zinc finger protein 9

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

EXPLORING THE MUSCLEBLIND SEQUESTRATION MODEL FOR RNA-MEDIATED DISEASES

By

Yuan Yuan

December 2007

Chair: Maurice S. Swanson
Major: Medical Sciences--Genetics

Nothing is static in living organisms. The life of an animal is spent adapting to ever-changing external and internal environments. Variations in the external environment can have profound effects on the behavior, fitness or even survival of an animal. Programmed development has been optimized during the course of evolution to increase animal survival rate under these variable external conditions. In mammals, metamorphosis happens in embryogenesis during which genes are switched on and off to initiate and regulate the formation of various organs and tissues. The end result of this regulation is a fully developed fetus prepared for the transition from the maternal to the outside environment. Programmed developmental events continue to take place postnatally to transform neonates to adult animals, inducing both qualitative and quantitative changes in tissues. However, the regulation of these processes is less understood compared to those regulating prenatal development. Myotonic dystrophy (DM) provides an excellent tool to study the fetal to adult transition. DM is an autosomal dominant neuromuscular disease caused by abnormal $(CTG)_n$ or $(CCTG)_n$ microsatellite expansions in the non-coding regions of the *DMPK* or *ZNF9* genes, respectively. A unique molecular feature of DM is that embryonic splicing patterns of certain pre-mRNAs persist in adult DM tissues, possibly causing the clinical features of this disease. Thus, the

mechanistic basis of DM is dysregulation of postnatal alternative splicing, and elucidating DM pathogenesis could uncover the associated factors and pathways vital for postnatal splicing switches. When transcribed, mutant *DMPK* and *ZNF9* microsatellites form stable RNA hairpins which sequester muscleblind-like (MBNL) proteins. Considerable data support the idea that DM is a MBNL loss-of-function disease. However, how MBNL sequestration caused DM was unclear. In this study, we characterized MBNL1 as one of the first alternative splicing factors which specifically function in developmentally regulated splicing.

The mechanistic basis for MBNL functional sequestration in DM was another unsolved puzzle in DM pathogenesis. The abilities of $(CUG)_n$ and $(CCUG)_n$ RNAs to titrate MBNL suggested that MBNL has a higher affinity for pathogenic RNAs versus its normal splicing targets. However, through a combination of RNA structural probing, filter binding and gel mobility shift assays, we demonstrated similar MBNL1 binding properties to pathogenic and splicing target RNAs, both in terms of binding affinities and recognition preferences. We also employed electron microscopy to show that MBNL1 forms oligomeric ring structures on CUG repeat RNA. While the amino-terminal region of MBNL1 is essential for RNA binding, the carboxyl-terminus mediates self-interaction which we hypothesize stabilizes inter-ring stacking. While most protein-RNA interactions are dynamic, we discovered that the interaction between MBNL1 and pathogenic RNAs, but not a normal splicing target, is unusually static. Based on this observation, we propose a model which explains how functional sequestration of MBNL is accomplished in DM.

CHAPTER 1
INTRODUCTION

Overview of Myotonic Dystrophy

Myotonic dystrophy (DM) is a multisystemic, autosomal dominant disease that affects 1 in 8,000 individuals worldwide (*1*). As perhaps the most variable of all human diseases, myotonic dystrophy can affect individuals of all ages ranging from infants to seniors. DM patients present an array of seemingly unrelated symptoms in a great variety of organs and systems including skeletal muscle, heart, brain, eye and the endocrine system (*1*). At least two types of myotonic dystrophies, myotonic dystrophy type I (DM1) and type II (DM2), have been characterized clinically and genetically in detail. Each of these DM types cause distinct yet similar clinical manifestations (*2*). The highly heterogeneous symptoms in DM1 and DM2 nonetheless reflect the unusual underlying genetic variations that have been pinpointed to two independent microsatellite expansions in the non-coding regions of two unrelated genes (*3-7*). Although DM was one of the first genetic anticipation disorders for which an unstable genetic element was identified, it was not until recently that the pathogenesis mechanism was unveiled (*8*).

Clinical Features of Myotonic Dystrophies

Myotonic dystrophy type I (DM1) - adult-onset

The hallmark of the DM1 skeletal muscle phenotype is myotonia, which can be elicited in almost every symptomatic adult DM1 patient, and also frequently found in presymptomatic individuals. Myotonia is defined as slow muscle relaxation after voluntary muscle contraction. In the clinic, it is readily tested in the hand muscles where, following a forceful contraction, there is a delayed ability to relax the grip (*1*). In DM1, myotonia is frequently accompanied by progressive muscle weakness and wasting in the facial muscles as well as the distal muscles in

the limbs. Skeletal muscle sections reveal an increased number of centralized nuclei and split fibers in DM1 muscle, presumably indicating immaturity of muscle fiber development (*1*).

Besides skeletal muscle involvement, DM1 patients present multisystemic manifestations including cardiac conduction defects, posterior iridescent ocular cataracts, irritable bowel syndrome, testicular atrophy, insulin resistance, hypogammaglobulinemia, hypersomnia and personality disturbance. Cardiac conduction defects probably account for the increased sudden death rate observed in DM patients (1,2,9-11). Notably, genetic anticipation, characterized by an intergenerational increase in disease severity and earlier age of onset, is striking in DM1. Mildly affected women may give birth to severely and congenitally affected children.

Myotonic dystrophy type I – congenital form (CDM)

Congenital DM1 presents a distinct clinical picture compared to adult-onset DM1. Affected infants are born with severe neurological, neuromuscular and musculoskeletal abnormalities including hypotonia (floppy baby), talipes, craniofacial abnormalities, poor suckling and swallowing and often require respiratory support to survive (*1*). Unlike adult-onset DM1, mental retardation is a common accompaniment and arguably the most important feature of CDM. Intriguingly, myotonia, the cardinal feature of adult-onset DM1, is never present in the first year of life in the CDM patients. However, CDM survivors in their adulthood eventually develop myotonia, among the full spectrum of adult-onset DM1 symptoms(1,9,10).

Myotonic dystrophy type II (DM2)

While DM1 can affect individuals of all ages, DM2 has a reported median age of onset at 48 years old(*12*). Similar to DM1, myotonia, progressive muscle wasting and weakness are the key features of skeletal muscle involvement in DM2. But the pattern of muscle weakness in DM2 is significantly different from that in DM1, in that DM2 patients frequently show proximal muscle weakness and wasting, such as in hip girdle muscles, as opposed to the distal muscle

weakness and wasting characteristic of DM1 patients. In addition, facial muscle weakness, which is prominent in DM1, is absent in DM2. Histologically, sections of DM2 muscles also reveal centralized nuclei as seen in DM1 muscles(1,9,10).

DM1 and DM2 share multisystemic effects in the eye, heart and endocrine system. Yet one of the primary differences between DM1 and DM2 is the lack of a DM2 form that affects development of muscles and the central nervous system. To date, there has not been any definitive evidence indicating that mental retardation is associated with DM2.

Figure 1-1 presents a brief comparison of the symptoms in DM1 and DM2.

Genetic Basis of Myotonic Dystrophies

Substantial progress in understanding the genetic basis of DM1 was made in 1992, when six groups mapped and further confirmed that the DM1 mutation was an unstable $(CTG)_n$ microsatellite expansion in the 3'-untranslated region (3'-UTR) of the dystrophia myotonica protein kinase (*DMPK*) gene (4-7,13,14). In addition, repeat size shows an inverse correlation with age of onset and a positive correlation with the severity of the disease, with 5-37 repeats in normal individuals, 50-1,000 in mild and classical adult-onset DM1 patients and >1,000 in the most severe cases, congenitally affected patients (*15*). Thus, the great genetic variability of the DM1 mutations is an important component of the extremely heterogeneous symptoms observed for DM1 patients. There is a marked tendency for expanded $(CTG)_n$ repeats to increase in size both in somatic tissues and in the germ line, in the latter case leading to more severe phenotypes in the offspring, thus accounting for the observed genetic anticipation in DM1 patients (12,16-22). Expansion of CTG repeats can be dramatic in DM1, especially through maternal transmission; in some cases expansions of several thousand repeats can occur in a single generation (22,23). Studies in various systems and organisms including cell culture, bacteria, yeast and mice, have shown that the CTG repeat instability involves DNA replication-dependent

as well as -independent mechanisms (*24-26*). Several DNA mismatch repair proteins such as Msh2, Msh3 and Pms2 have been shown to play important roles in somatic and germline expansion of the repeats (*27-32*). Genetic anticipation has also been observed in a number of other diseases including Huntington disease (HD), spinocerebellar ataxias (SCA), schizophrenia and bipolar disorders. Interestingly, some, if not all, of these diseases are also caused by unstable repeat expansions (*33-38*).

Similar to DM1, DM2 is caused by another unstable DNA element, a $(CCTG)_n$ tetranucleotide expansion located in the first intron of zinc finger protein 9 (*ZNF9*) (*3*). *ZNF9* encodes a nucleic acid binding protein that is highly expressed in skeletal muscle and has been suggested to modulate the activity of beta-myosin heavy chain and more recently, internal ribosomal entry site (IRES) mediated translation (39,40). Positioned at ~850 nt upstream of the 3' splice site of *ZNF9* intron1, the repeat expansion has been shown not to affect ZNF9 mRNA processing or protein expression (*41*). While *ZNF9* $(CCTG)_n$ repeats in normal individuals are short (5-26) and often interrupted by other sequences, the CCTG portion expands in affected individuals to ~75-11,000 repeats (*42*). Interestingly, although DM2 is generally a milder disease compared to DM1, the average size of the DM2 CCTG expansion is around 5,000, bigger than the longest documented DM1 CTG expansion (~4,000 repeats). Unlike DM1, there is a lack of positive correlation between the size of the repeats and the severity of the disease. Reports have shown that even people homozygous for large $(CCTG)_n$ expansions do not show more severe symptoms(12,43). In addition, there is no convincing evidence for genetic anticipation in DM2, although the CCTG tetranucleotide expansions do display a high degree of instability in both somatic and germline cells (*42*).

Models for DM Pathogenesis

For years after the discovery of the DM1 and DM2 mutations, one in the 3'-UTR of the *DMPK* gene and the other in the first intron of *ZNF9*, a fundamental question about the pathogenic mechanism remained unanswered. How do microsatellite expansions in these non-coding regions cause dominantly inherited diseases? At least two types of models, which are not necessarily mutually exclusive, have been proposed for DM pathogenesis.

Haploinsufficiency models

Since the $(CTG)_n$ expansion resides in the *DMPK* gene, the DMPK haploinsufficiency model was initially proposed for DM1 pathogenesis. This model hypothesizes that a reduction of DMPK protein levels caused by the microsatellite expansion is responsible for the multisystemic clinical features of DM1. Indeed there is evidence supporting reduced DMPK levels in adult DM1 patients (*44-46*). However DMPK levels in congenital DM1 patients appear to be unaltered compared to normal individuals (*47*). Furthermore, *Dmpk* heterozygous or homozygous knockout mice fail to recapitulate the cardinal phenotypes of human DM1 patients. Instead, these mice develop a cardiac conduction defect and a mild, late-onset myopathy which is not characteristic of DM1 (*48-50*). Taken together, these results suggest that although DMPK haploinsuffciency in adult DM1 patients may contribute to cardiac features of DM1, loss-of-function of DMPK is not the sole requirement for the multisystemic progression of the disease.

A second kind of haploinsufficiency model proposes that $(CTG)_n$ repeats constitute a strong nucleosome positioning sequence, which could alter local chromatin structure and lead to reduced expression of neighboring genes, dyotrophia myotonica-containing WD repeat motif (*DMWD*) and sine oculis-related homeobox 5 homolog (*SIX5*) (51,52). Interestingly, *DMPK* gene is located in a gene rich cluster where the *DMWD*, *DMPK* and *SIX5* genes occupy only a 20 kilobase (kb) region. The 3' terminus of the *DMWD* gene overlaps the *DMPK* promoter and the

CTG expansion overlaps the 5' promoter region of *SIX5*. Although experiments designed to test for *DMWD* and *SIX5* expression in DM1 patients have so far yielded inconsistent results (*53-56*), the SIX5 haploinsufficiency model gained support from studies showing that *Six5* heterozygous and homozygous knockout mice develop ocular cataracts (57,58). However, these cataracts are not the iridescent opacities with posterior location commonly seen in DM patients.

Of course, the DMPK and SIX5 haploinsufficiency models only provide an explanation for DM1, and not DM2, pathogenesis. For DM2, neither *ZNF9* gene expression nor the processing of ZNF9 mRNA is affected by the CCTG expansion (*41*). The high degree of clinical parallels between DM1 and DM2 suggest very similar underlying molecular events leading to these disorders.

RNA dominance models

In 1995, the Singer and Housman labs discovered that RNA transcripts containing either expanded CUG or CCUG repeats accumulate in nuclear foci both in cultured cells and biopsied tissues (*59*). This was the first indication that mutant RNAs are directly involved in DM pathogenesis. In 1996, Timchenko and colleagues proposed the RNA dominance model which hypothesized that mutant RNAs transcribed from the DM loci exert a toxic gain-of-function effect, possibly by disrupting other cellular processes such as transcription and RNA processing (*60*). The strongest support for the RNA dominance model comes from a transgenic mouse model developed in 2000. Mankodi and colleagues generated transgenic mouse lines expressing either 5 or 250 CTG repeats inserted at the 3'-UTR of the human skeletal muscle actin gene which is not implicated in DM pathogenesis and expressed only in skeletal muscle. While the mice expressing $(CTG)_5$ were normal, mice expressing the longer CTG repeats developed ribonuclear inclusions and displayed myotonia and myopathic features characteristic of DM1 muscle (*61*). Furthermore, the toxic effect of the transgene is dependent on its expression level.

Mice with transgene insertions in heterochromatic regions appeared phenotypically normal, indicating a trans-dominant effect at the RNA, instead of the DNA, level.

Yet questions remain as to how the repeat-containing RNAs elicit toxic effects. The RNA structure prediction program Mfold suggests that CUG and CCUG repeats form hairpin structures *in vitro* with G-C basepairs and mismatched pyrimidines (*62*). This prediction was later supported by crystallography and confirmed by RNA structural probing and electron microscopy (*63-67*). The length of these hairpins is proportional to the number of repeats. Thus, cellular factors with high affinities for such hairpin structures could potentially be titrated by these abnormal RNAs away from their physiological targets. In an attempt to isolate such factors, Miller and colleagues identified human muscleblind-like (MBNL) proteins as strong candidates (*68*). MBNL proteins bind to CUG RNA hairpins in a length-dependent manner. Moreover, MBNL proteins colocalize with CUG or CCUG containing RNAs in ribonuclear inclusions in DM myoblasts as well as tissues. These observations lead to the MBNL sequestration model, which proposes that *DMPK* and *ZNF9* mutant RNAs disrupt normal MBNL function, and thus DM is a MBNL loss-of-function disease. To test for this hypothesis, Kanadia and colleagues generated an *Mbnl1* knockout mouse model which will be discussed further in the following chapter. Consistent with the MBNL sequestration model, Mbnl1 loss-of-function in the mouse elicits key DM features, including myotonia, structural changes in muscle along with posterior "dust-like" cataracts (*69*). This result strongly argues that titration of MBNL proteins underlies DM pathogenesis.

Although the MBNL sequestration model for DM pathogenesis has gained considerable experimental support to date, alternative hypotheses have been raised, including transcription factor sequestration proposed by Ebralidze and colleague (*70*). Upon expression of CTG repeat-

containing DMPK RNA in myocytes, certain transcription factors, such as Sp1and retinoic acid receptor gamma (RARγ), relocalize from chromatin to the ribonuclear inclusions. Subsequent reduction in the mRNA levels was detected for skeletal muscle chloride channel 1 (*CLCN1*), which harbors multiple Sp1 binding sites in its promoter region. Replenishment of Sp1 in mutant RNA expressing cells partially restored the *CLCN1* mRNA level. Based on these data, Ebralidze and coworkers postulated that leaching of Sp1 by mutant RNAs contributes to reduced muscle chloride channel 1 levels, and resulted in myotonia in DM1 patients (*70*).

In summary, our knowledge on DM pathogenesis has rapidly expanded over the past decade. The prevailing RNA dominance view is backed up by numerous *in vitro* and *in vivo* studies. Sequestration of human muscleblind-like proteins by expanded CTG or CCTG repeats may be the major mechanism underlying RNA toxicity.

Human Muscleblind-Like Proteins

Structure and Phylogeny

Based on considerable sequence homology, human muscleblind-like proteins were named after their *Drosophila* homolog, muscleblind (*mbl*), which, when mutated, results in abnormalities in the terminal differentiation of muscle and photoreceptor cells (71,72). To date, muscleblind proteins have been found in a variety of metazoan phyla, including nematoda, arthropoda, tunicata and vertebrata. While a single muscleblind gene was found for each invertebrate species, most vertebrates have three, and in some case, such as *Takifugu rubripes,* up to five muscleblind homologs (*73*). Human *MBNL* family includes three members, *MBNL1, 2* and *3*, located on chromosomes 3, 13 and X chromosome, respectively. Similar genomic organizations of the three human muscleblind-like genes suggest that they originated from gene duplication events during evolution.

The most distinct structural feature of muscleblind proteins is the presence of tandem CCCH zinc finger motifs, composed of three cysteine and one histidine. Known for its ability to mediate nucleic acid-protein interactions, the CCCH motif is found in many RNA and DNA-binding proteins, including mouse tristetraprolin (TTP) and yeast nuclear polyadenylated-RNA binding protein 2 (Nab2p) (74,75). The structure of an RNA-bound CCCH containing protein, zinc finger protein 36 C3H type-like 2 (ZNF36L2), has been solved using nuclear magnetic resonance (NMR). This structure indicates that electrostatic as well as hydrogen-bonding interactions between closely stacked RNA bases and protein aromatic side chains play important roles in stabilizing the protein-RNA interaction (*76*). While all muscleblind proteins contain CCCH motifs, the number of the motifs in each protein varies from nematoda to vertebrata. Protostome species encode musclebind with only two CCCH motifs, whereas four such motifs are present in muscleblind proteins from Tunicata and Vertebrata species, arranged as two pairs separated by a linker region of around 60 amino acid residues (*73*). Comparison of the individual CCCH motifs amino acid compositions in *Drosophila* and human muscleblind proteins reveals that the first and third CCCH motif in human muscleblind proteins resemble the first CCCH motifs in *Drosophila* protein, while the second and fourth CCCH motifs in human resemble the second CCCH motif in *Drosophila*. This suggests that the CCCH motifs duplicate, and function as pairs during evolution.

Muscleblind proteins show strong functional conservation among species. Although human and *Drosophila* muscleblind proteins display a distant 31% homology, human MBNL proteins can nevertheless rescue *Drosophila mbl* mutants (*77*). Among vertebrate species, muscleblind proteins are also highly conserved in sequence. Mouse and human MBNL1 differ by only two amino acid residues. Interestingly, the high degree of conservation is found not only at

the protein level, but also at certain genomic regions. As shown by Bejerano and colleagues, *MBNL1* and *MBNL2* are both among a group of genes which have an uninterrupted >200 base pair (bp) genomic segments that are completely identical in human, mouse and rat. These so called "ultraconserved regions" in *MBNL1* and *MBNL2* overlap each other and are located around exon 7 of *MBNL1* and the corresponding exon 6 of *MBNL2* (*78*). Such strong conservation could be a result of potential mutation cold-spots, although the fact MBNL1 and MBNL2 are located in two distinct chromosomal loci would argue against this possibility. Alternatively, extreme negative selection, which usually indicates functional importance, could be responsible for the ultraconservation. The mechanism for this absolute sequence conversation in human, mouse and rat is not yet clear.

Expression Profile

Human muscleblind-like family members show distinct expression patterns in adults. By Northern blot analysis, *MBNL1* is abundantly expressed in heart and skeletal muscle, while *MBNL2* shows a more ubiquitous expression pattern with its mRNA detected in virtually every tissue examined. In contrast to *MBNL2*, a very specific pattern was seen in the case of *MBNL3*, which is expressed most strongly in placenta while showing very weak signal in pancreas, kidney, liver and heart (68,79).

The embryonic expression patterns of human *MBNL* genes, however, have to be extrapolated from the data obtained in mice (*80*). Using RNAs derived from embryos of different stages, RNA blot analysis revealed coordinate temporal expression patterns of *Mbnl*, with *Mbnl1* peaking between 13.5-15.5 days post coitus (dpc), *Mbnl2* between 17.5-18.5 dpc and *Mbnl3* between 11.5-15.5 dpc. Furthermore, whole-mount in situ hybridization was performed on 9.5 dpc embryos and all three muscleblind proteins were highly expressed in the developing head region with prominent *Mbnl2* and *Mbnl3* expression in the mandibular and maxillary

components in the first branchial arch and neural tube. All three *Mbnl* genes were moderately expressed in forelimb buds. Interestingly, *Mbnl2* also showed strong expression in the heart. When examined at a slightly later stage, *Mbnl1* and *Dmpk* expression patterns overlap in the hindlimb and forelimb buds as well as the first and second branchial arches. This observation supports the RNA dominance model, indicating that *Dmpk* mutant transcripts could titrate Mbnl1 during embryonic development and result in the developmental defects seen in CDM patients (*80*).

Cumulatively, these results suggest that myotonic dystrophy is caused by mutant RNAs which, in turn, lead to sequestration of MBNL proteins. However, an important question remains. What is the missing link between MBNL loss-of-function and DM phenotypes? In other words, what is the physiological function(s) of MBNL proteins?

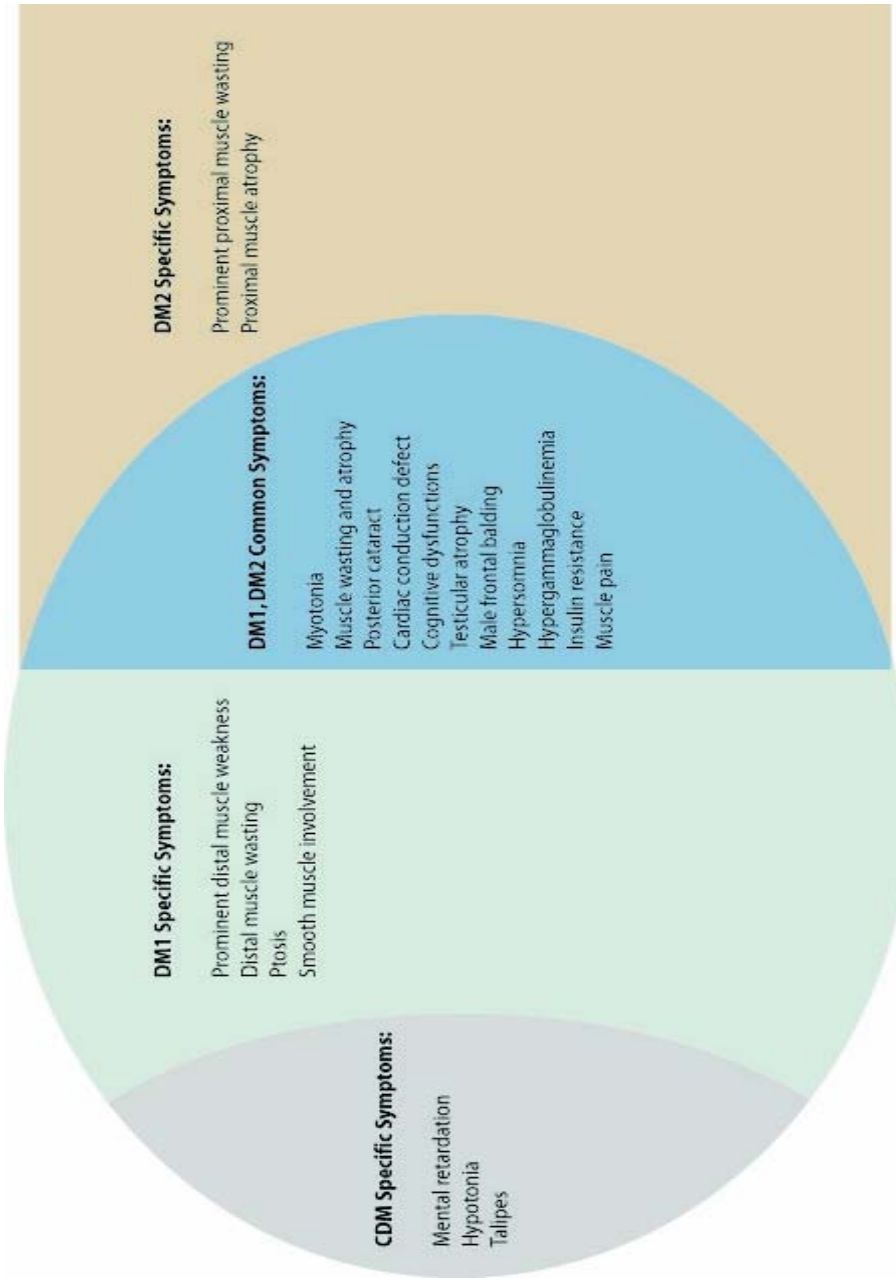

Figure 1-1 Comparison of adult-onset DM1, congenital DM1 and DM2 symptoms.

CHAPTER 2
MATERIAL AND METHODS

Plasmids and Constructs

To construct pGEX-6P-His, a *Xho*I-*Not*I fragment encoding the His_6 tag from pGEX-MBNL1 (NP_066368) was inserted into *Xho*I-*Not*I digested pGEX-6P-1 (GE Life Sciences). pGEX-6P-MBNL1-His was constructed by inserting a *Bam*HI-*Xho*I fragment from pGEX-MBNL1 into *Bam*HI-*Xho*I digested pGEX-6P-His. The MBNL1 N-terminal region (residues 1-253) was amplified using primers MSS2759 and MSS2760 (see Appendix B), digested with *Bam*HI and *Xho*I and inserted into pGEX-6P-His to create pGEX-6P-MBNL1-N-His. For the yeast two hybrid system, either full length, amino terminal (residues 1-264) or carboxyl terminal (residues 239-382) MBNL1 cDNAs were PCR amplified using the following primers and inserted into pGBKT7 or pGADT7 (Clontech, Mountain View, CA) at *Sma*I and *Bam*HI sites: 1) full length, primers MSS1163 (forward primer) and MSS1166 (reverse primer); 2) N-terminal region, MSS1164 (forward) and MSS1166 (reverse); 3) C-terminal region, primers MSS1163 (forward) and MSS1165 (reverse). To create pcDNA3-V5, primers MSS3045 and MSS3046 were subjected to a 10-cycle PCR reaction: 94^{o}C 30 sec, 50^{o}C 20 sec, 72^{o}C 20 sec. The resulting DNA fragment was gel purified followed by digestion with *Nhe*I and *Bam*HI, and inserted into *Nhe*I-*Bam*HI digested pcDNA3.1(+) (Invitrogen, Carlsbad, CA). The *Bam*HI-*Xho*I fragment from pGEX-6P-MBNL1-His was inserted into pcDNA3-V5 at *Bam*HI and *Xho*I sites to create pcDNA-V5-MBNL1. The *Bam*HI-*Xho*I fragments from pGEX-6P-MBNL1-His and pGEX-6P-MBNL1-N-His were inserted into pcDNA3.1(+)/myc-His A (Invitrogen) to create pcDNA3-MBNL1-mycHis and pcDNA3-MBNL1-N-mycHis, respectively. The CUGBP1 (NP_006551) coding sequence was PCR amplified using primers MSS2699 and MSS2700 and inserted into pcDNA3.1(+)/myc-His A at *Bam*HI and *Xho*I sites. To construct pEGFP-C1-hnRNP A1

(NP_034577), the mouse hnRNP A1 coding region was amplified using primers MSS2002 and MSS2003. PCR product was digested with EcoRI and XhoI, and inserted into EcoRI-XhoI digested pEGFP-C1(Invitrogen). For pEGFP-C1- MBNL1-N, the MBNL1 N-terminal coding region (residues 1-253) was amplified using primers MSS3074 and MSS3075, digested with XhoI and BamHI and inserted into pEGFP-C1. The Tnnt3 minigene was prepared by amplifying the mouse *Tnnt3* genomic region between exons 8 and 9 using primers MSS1949 and MSS1950 and inserting the PCR product into pSG5 (Stratagene, La Jolla, CA) at the *Eco*RI site. The mutant Tnnt3 minigenes, pSG5-Tnnt3Δ10 and pSG-Tnnt3/gg, were generated by site-directed mutatgenesis using primers MSS2129/MSS2130 and MSS3131/3132, respectively. Wild type pSG5-Tnnt3 (100 ng) (with 125 ng of each of the primers) was subjected to the following PCR reaction: 94^{o}C 30 sec, 50^{o}C 1 min, 72^{o}C 8 min, 20 cycles using Pfu DNA polymerase (Stratagene). After *Dpn*I digestion, the PCR product was transformed into DH10B and mutants were identified by plasmid DNA sequencing. Using pSG5-Tnnt3 as a template, PCR fragments generated from primer pairs MSS1865/MSS1879 and MSS1884/MSS1866 were TOPO-cloned into pCR4-TOPO (Invitrogen) to make pTOPO-T5.1 and pTOPO-T5.45, respectively.

Recombinant Protein Preparation

For the preparation of recombinant proteins, BL21(DE3) RP containing pGEX-6P-1-MBNL1 or pGEX-6P-1-MBNL1-N were grown to OD_{600}=0.5 followed by induction with 1 mM IPTG for 2 h at 30^{o}C. Cells were collected and resuspended in lysis buffer containing 25 mM Tris-Cl, pH 8.0, 0.5 M NaCl, 10 mM imidazole, 2 mM β-mercaptoethanol, 2 mg/ml lysozyme, 10 μg/ml DNase I, 5% glycerol, 0.1% Triton X-100 supplemented with protease inhibitors. The cell suspension was incubated on ice for 30 min with stirring prior to sonication and centrifugation at 12,000 X *g*. For protein purification, Ni-NTA-Sepharose (Amersham/GE

Healthcare, Piscataway, NJ) (12 ml) was incubated with the supernatant for 1 h at 4°C and washed three times with 40 ml of wash buffer containing 25 mM Tris-Cl, pH 8.0, 0.5 M NaCl, 20 mM imidazole, 0.1% Triton X-100, followed by three 10 ml elutions in 25 mM Tris-Cl, pH 8.0, 0.5 M NaCl, 250 mM imidazole, 0.1% Triton X-100. Subsequently, β-mercaptoethanol was added (10 mM final concentration) to the eluate which was incubated with 2 ml glutathione-Sepharose (Amersham) for 1 h at 4°C. After three washes (10 ml each) of buffer (WB) containing 25 mM Tris-Cl, pH 8.0, 300 mM NaCl, 5 mM β-mercaptoethanol, 0.1% Triton X-100, the glutathione-Sepharose beads were incubated with 4 ml WB containing 40 U of PreScission protease (Amersham) at 4°C overnight. The supernatant was collected following brief centrifugation and concentrated to 1-8 mg/ml).

Chemical and Enzymatic Analysis of RNA Structures

Transcription reactions were carried out in a 50 μl volume which contained 2 μg of each DNA template, 1 mM NTPs, 3.3 mM guanosine, 60 units of ribonuclease inhibitor RNase Out (Invitrogen), 200 units of T7 RNA polymerase (Ambion, Austin, TX), 10 mM DTT, 40 mM Tris-HCl (pH 7.9), 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl. The reaction was performed at 37°C for 2 h, the transcript was then purified on a denaturing 10% polyacrylamide gel and subsequently 5'-end-labelled with T4 polynucleotide kinase and [γ^{32}P]ATP (3000 Ci/mmol). The labeled RNA re-purified by electrophoresis on a denaturing 10% polyacrylamide gel. Prior to structure probing, the labeled RNA was subjected to a denaturation/renaturation procedure in a reaction buffer containing 50 mM Tris–HCl (pH 8.0), 60 mM KCl, 15 mM NaCl, 2 mM $MgCl_2$ by heating the sample at 90°C for 1 min. and slowly cooling to 25°C. The RNA sample was then mixed with either a 25-fold molar excess of recombinant MBNL1 in 50 mM Tris–HCl pH 8.0, 60 mM KCl, 15 mM NaCl, 2 mM $MgCl_2$, 2% glycerol, 0,5 mM DTT, 50 μg/mL BSA, or with

buffer only (control) and incubated 20 min at 25°C. The final concentration of $(CUG)_{54}$ was 20 nM and MBNL1 was 500 nM. Under these reaction conditions more >95% of RNA was bound with protein as revealed by filter binding assays. Additional control samples were prepared by mixing RNA with MBNL1 protein previously denatured by heating at 75°C for 2 min. Limited RNA digestion was initiated by mixing 5 µl of the RNA or RNA/protein sample (25,000 cpm) with 5 µl of a probe solution containing either lead ions or ribonuclease T1 in reaction buffer. The reactions were performed at 25°C for 20 min. and stopped by adding 20 volumes of 1XTE buffer followed by phenol/chloroform extraction. Precipitated RNAs were dissolved in a denaturation solution (7.5 M urea and 20 mM EDTA with dyes). To determine the cleavage sites, the products of RNA fragmentation were separated on 10% polyacrylamide gels containing 7.5 M urea, 90 mM Tris-borate buffer, and 2 mM EDTA, along with the products of alkaline hydrolysis and limited T1 nuclease digestion of the same RNA. The alkaline hydrolysis ladder was generated by the incubation of the labeled RNA in formamide containing 0.5 mM $MgCl_2$ at 100°C for 10 min. The partial T1 ribonuclease digestion of RNAs was performed under semi-denaturing conditions (10 mM sodium citrate, pH 5.0; 3.5 M urea) with 0.2 unit/µl of the enzyme during incubation at 55°C for 10 min. Electrophoresis was performed at 1800 V (gel dimensions, 30/50 cm). The products of the structure probing reactions were visualized by PhosphorImaging (Storm; Molecular Dynamics, Sunnyvale, CA) and analyzed by ImageQuant 5.2 (Molecular Dynamics).

Isolation of RNA from mouse skeletal muscle and RT-PCR

Quadriceps isolated from 129 mice at different post-natal ages were each homogenized in 0.5 ml of TRI reagent (Sigma, T9424) at 4°C. Total RNA was isolated according to manufacturer's instructions. The RNA was further treated with RQ1 DNase I followed by phenol extraction and ethanol precipitation. Reverse transcription was performed according to

manufacturer's instructions in a 20 μl volume containing 1 X first strand buffer, 10 mM DTT, 5 μg of RNA, 300 ng of random hexamers (Invitrogen), 0.5 mM dNTPs and 200 U of Superscript II RNase H^- reverse transcriptase (Invitrogen). PCR was performed in a 50 μl reaction volume using 4 μl of cDNA, 30 pmol of each primer (Primer1: TCTGACGAGGAAACTGAACAAG, Primer 2: TGTCAATGAGGGCTTGGAG), 10 μCi of α-^{32}P dCTP and 2.5 U of Taq (Invitrogen) for 27 cycles of 94^{o}C for 30 sec, 50^{o}C for 30 sec, 72^{o}C for 30 sec, followed by one cycle at 72^{o}C for 7 min. PCR samples were purified using a Qiagen PCR purification kit. Half of the eluate was digested with BsrBI and the digested and undigested PCR products were separated on a 8% polyacrylamide gel prior to autoradiography.

Cotransfection Assays

HEK293T or C2C12 cells were plated in a 6 well plate in antibiotic-free medium to 30% confluency. About18-24 hr later, cells were transfected with 1 μg of minigene and 2 μg protein expression plasmid using Lipofectamine 2000 (Invitrogen) following manufacturer's protocol. Forty-eight hours post-transfection, cells were lysed in 0.2 ml of TRI reagent (Sigma), and total RNA was isolated following the manufacturer's protocol. Reverse transcription was performed as described above except that oilgo(dT) 12-18 was used instead of random hexamers. The PCR reactions were performed using 4 μl of cDNA and primers flanking pSG5 MCS (Primer1: AGAATTGTAATACGACTCACTATAGGGC, Primer 2: GCTGCAATAAACAAGTTCTGCTTT) for 27 cycles of 94^{o}C for 30 sec, 55^{o}C for 30 sec and 72^{o}C for 20 sec. Fifteen microliters of PCR reactions were separated on a 15% polyacrylamide gel at 75 volts for 16 hr followed by autoradiography.

Yeast Two-hybrid Analysis

pGBKT7-MBNL1 together with one of the pGADT7 constructs (pGADT7-T Antigen, pGADT7-MBNL1, pGADT7-MBNL11-N, pGADT7-MBNL1-C) were transformed into yeast strain AH109. In addition, pGBKT7-p53 and pGADT7-T Antigen (Clontech) were co-transformed into AH109 as a negative control. Double transformants were selected on SD/-Trp/-Leu plates. Expression of the myc-tagged GAL4 DNA binding domain (BD) and HA-tagged GAL4 activation domain (AD) fusion proteins were confirmed by immunoblot analysis using mAb 9E10 and 16B12, respectively. To test for protein interactions, transformants were streaked onto SD/-Trp/-Leu/-His incubated at 30^{o}C for 3-4 days and scored for growth.

Co-immunoprecipitation

To test for MBNL1-MBNL1 interactions in mammalian cells, HEK293T cells were transfected with 5 μg of pcDNA3-V5-MBNL1 alone as a control or 5 μg of pcDNA3-V5-MBNL1 together with either 5 μg of pcDNA3-MBNL1mycHis or pcDNA3-MBNL1-N-mycHis. Cells were harvested 20-24 h post-transfection by trypsinization followed by neutralization in media contain 10% fetal bovine serum and two washes in 50 mM Tris-HCl, pH7.4, 150 mM NaCl. Cell pellets were resuspended in 50 mM Tris-HCl, pH7.4, 150 mM NaCl, 0.1% IGEPAL with protease inhibitors, and sonicated on ice (3 X 5 sec). Cell debris was removed by centrifugation at 16,100 x *g* for 10 min at 4^{o}C. Cleared lysates were treated with 200 μg/mL RNase A for 20 min on ice (17) followed by another 10-minute centrifugation. Cleared lysates were mixed with Dynabeads coupled to Protein A (Invitrogen) precoated with rabbit anti-V5 polyclonal antibody (Novus, Littleton, CO) and incubated at 4^{o}C for 2 hr. Dynabeads were washed three times with IPP150 buffer (50 mM Tris-HCl, pH7.4, 150 mM NaCl, 0.1% IGEPAL) and once with 50 mM Tris-HCl, pH7.4, 150 mM NaCl. Proteins were dissociated from the beads

by heating at 95°C for 2 min in 1X SDS-PAGE sample buffer. Proteins (50% of the immunoprecipitated proteins, 2.5% of the input) were separated on 12.5% SDS-PAGE gels. Immunoblotting was performed using mAb 9E10 (1:1000) or mAb anti-V5 (1:1000, AbD Serotec).

Electron Microscopy

The CUG_{136} RNA, transcribed from pBC-CTG136 plasmid (R. Osborne, University of Rochester), was incubated with MBNL1 protein at molar ratio of 1:2.5 or 1:10 in a buffer containing 16 mM HEPES, 2 mM magnesium acetate, 0.16 mM EDTA, 0.4 mM DTT, 1 mM ATP, 50 mM potassium acetate, and 16% glycerol for 30 min at 30°C. The resulting complexes were fixed in 0.6% glutaraldehyde for 5 min at room temperature and subsequently passed over a 2 ml column containing Bio-Gel A-5M (Bio-Rad, Hercules, CA) equilibrated with 0.01 M Tris (pH 7.6) and 0.1 mM EDTA to remove free protein and fixatives. Protein-RNA enriched fractions were incubated with 2.5 mM spermidine and adsorbed on to glow charged carbon coated copper grids and dehydrated in a series of ethanol washes of 25, 50, 75, and 100% ethanol each for 5 min at room temperature. Samples were air dried prior to rotary shadow casting with tungsten. Protein-RNA complexes were visualized using a FEI Tecnai 12 electron microscope (FEI, Hillsboro, OR) at an accelerating voltage of 40 kV and images were captured using a 4k x 4K Gatan CCD camera using plate film or Gatan digital image capturing software (Gatan, Pleasanton, CA). Plate film negatives were scanned using an Imacon scanner and supporting software (Imacon, Redmond, WA). Images were photographed at a magnification of 52K.

Photocrosslinking

To prepare whole cell lysates for crosslinking, HEK293T cells were grown in 10 cm plates and transfected with 10 μg of pcDNA3-CUGBP1mycHis, pcDNA3-MBNL1mycHis or

pcDNA3-MBNL1-NmycHis. Cells were trypsinized and then neutralized in media containing 10% FBS followed by two additional PBS washes. Cells were then resuspended in 250 μL of 20 mM HEPES-KOH (pH 8.0), 100 mM KCl, 0.1% IGEPAL and protease inhibitors, sonicated, and lysates centrifuged (13,200 rpm, 10 min, 4^{o}C). Glycerol was added to the supernatants to a final concentration of 20%.

RNAs for photocrosslinking were uniformly labeled with 40 μCi each of (α-^{32}P)-GTP and (α-^{32}P)-UTP (800 Ci/mMole) in the presence of 0.5 mM ATP and CTP, 0.02 mM GTP and UTP. Crosslinking was performed by incubating 0.1 pmol RNA with 15 μL of HEK293T whole cell lysate in 25 μl reactions containing 16 mM HEPES-KOH (pH 8.0), 65 mM potassium glutamate, 2 mM $Mg(OAC)_2$, 0.4 mM DTT, 0.16 mM EDTA, 20 mM creatine phosphate, 2 mM ATP and 16% glycerol (final concentration). Reactions were incubated at 30^{o}C for 15 min, transferred to pre-chilled PCR caps on ice and photocrosslinked in Stratalinker (Stratagene, La Jolla, CA) for 2.5 min (three times) with a 3 min interval between each irradiation. Samples were digested with 5 μg of RNase A for 20 min at 37^{o}C and immunopurified using the anti-myc monoclonal antibody 9E10 pre-coated protein A Sepharose (Amersham). Purified proteins were fractionated on 12.5% polyacrylamide gels containing SDS (SDS-PAGE) followed by autoradiography.

A few modifications were made for the commitment assay. Cold RNA competitors were synthesized using the Megashortscript T7 kit (Ambion), followed by purification using the Megaclear kit (Ambion) according to manufacturer's instructions. Unlabeled RNAs were resuspended to 20 μM in final crosslinking buffer containing 16 mM HEPES-KOH (pH 8.0), 60 mM KCl, 65 mM potassium glutamate, 2 mM magnesium acetate, 0.4 mM DTT, 0.16 mM EDTA, 20 mM creatine phosphate, 2 mM ATP and 16% glycerol. Ten μl of unlabeled RNAs (200 pmol) were added to 15 μl of HEK293T whole cell lysate either at the same time with 0.1

pmol radiolabeled RNAs and followed by 1 hr incubation at 30°C, or alternatively 15 min after the incubation of HEK293T whole cell lysate with radiolabeled RNAs, and followed by a 45-min incubation at 30°C. Reactions were photocrosslinked as described above and treated with 20 μg of RNase A for 20 min at 37°C. The following immunoprecipitation and SDS-PAGE fractionation were performed as described above.

Filter-Binding Assays

Uniformly labeled RNA was prepared as described previously (*68*). Calibration of the non-specific retention rate of the nitrocellulose filter was performed by incubating 0.01-0.1 nM RNA at 30°C for 30 min in buffer containing 50 mM Tris-HCl (pH8.0), 40 mM KCl, 20 mM potassium glutamate, 15 mM NaCl, 0.5 mM DTT, 0.05 U/μL SUPER-asin (Ambion) followed by filtration through a Bio-Dot (BioRad) apparatus containing a sandwich of nitrocellulose (BioRad) and Hybond-N plus (Amersham) membranes followed by a single wash step with the same buffer. The membranes were UV-crosslinked, air dried and exposed to a phosphorimager screen. Non-specific retention on the nitrocellulose membrane was undetectable. Binding reactions were set up in the same buffer with 5 pM RNA and 3.13 X 10^{-12} M to 1.02 X 10^{-7} M of MBNL1/41-His, and incubated at 30°C for 30 min. Each reaction was applied to the Bio-Dot apparatus followed by one wash with binding buffer. Membranes were processed as described above and signals quantified using ImageQuant TL (Amersham). Standard deviations were calculated based on three independent experiments and apparent dissociation constants were calculated using a one-site binding model and GraphPad Prism (v3.00) software.

Electrophoretic Mobility Shift Assays

RNA was uniformly labeled with 40 μCi (α-^{32}P)-GTP or UTP (800 Ci/mMole) in the presence of 0.5 mM ATP, 0.5 mM CTP, 0.02 mM GTP, 0.02 mM UTP and purified using a 5%

denaturing gel containing 8 M urea. Prior to use, purified RNA was heated at 65^{o}C for 5 min in 50 mM Tris-HCl (pH8.0), 40 mM KCl, 20 mM potassium glutamate, 15 mM NaCl, 0.5 mM DTT, 0.5 U/μL SUPER-asin (Ambion) following by renaturation at room temperature. Reactions (20 μL) were assembled with 0.1 nM RNA and 0-256 nM protein in 50 mM Tris-HCl, pH 8.0, 40 mM KCl, 20 mM KGlutamate, 15 mM NaCl, 15% glycerol, 0.5 mM DTT, 20 μg/mL acetylated BSA. After incubation at 30^{o}C for 30 min, reactions were immediately loaded onto a 4% polyacrylamide gel (80:1) containing 0.5 mM DTT and 5% glycerol which had been pre-run at 150V for 1-2 h at 4^{o}C. Gels were run in 0.5X TBE (pH 8.3) at 200V for 2 h, fixed and dried prior to autoradiography.

CHAPTER 3
HUMAN MUSCLEBLIND-LIKE 1 PROTEIN REGULATES DEVELOPMENTAL SPLICING SWITCHES

Introduction: Misregulation of Alternative Pre-mRNA Splicing in Myotonic Dystrophy and Other Human Diseases

Overview of Alternative Splicing

Alternative splicing is a co-transcriptional process by which different forms of mRNAs are generated from the same gene. At least 70% of human genes undergo alternative splicing, which allows the Human Proteome Initiative (HPI) estimated 1 X 10^6 human proteins to be encoded by merely 2-2.5 X1 0^4 protein-encoding genes (81). There are five types of alternative splicing (Figure 3-1), with cassette exon splicing being the most common type (82). Alternative splicing provides an important mechanism for fine-tuning protein functions. While protein isoforms usually share long stretches of sequence in common, isoforms encoded by alternatively spliced mRNA species contain distinct sequences which potentially enable them to carry out slightly, or even drastically, different tasks in the cell. Intricate regulation of alternative splicing can often be observed in a cell-type or developmental-stage specific manner. When this regulation goes astray, detrimental effects can occur and lead to various human diseases.

Alternative Splicing Regulation

The general mechanism of pre-mRNA splicing must be introduced before discussing alternative splicing regulation. An intron of any pre-mRNA is defined by three major *cis*-regulating elements, the 5' splice site, the 3' splice site and the branch point. All pre-mRNA splicing events follow an identical two-step process (Figure 3-2). In the first step, the branch point adenosine carries out a nucleophilic attack on the 5' splice site, forms a 5'-2' phosphodiester bond with the nucleotide at the 5' end of the intron, which results in pre-mRNA

cleavage. In the second step, the newly created 3'-OH group of the 5' exon attacks the 3' splice site, which leads to ligation of the two exons and release of the intron lariat (83).

Despite this chemical simplicity, these two transesterification reactions are carried out by a complicated and dynamic mega-dalton RNA-protein complex called the spliceosome. The U2-dependent spliceosome, also known as the major spliceosome, which carries out >99% of splicing in the cell, is assembled from U1, U2, U4/U5/U6 small nuclear ribonucleoproteins (snRNPs) and numerous auxiliary protein factors which associate and dissociate with the spliceosome at different stages. During splicing, five spliceosome intermediates, E, A, B, B* and C complexes, have been observed (Figure 3-2). U1 snRNP is first recruited to the 5' splice site through base pairing with the 5' splice site, forming the spliceosomal E complex (Figure 3-2). Then with the help of the U2 auxiliary factor (U2AF), a heterodimer composed of 35 and 65 kDa subunits, U2 snRNP stably associates with the branch point sequence (BPS), thus forming the A complex. Subsequent recruitment of U4/U5/U6 tri-snRNP generates the pre-catalytic B complex. Major rearrangements of RNA-RNA and RNA-protein interactions occur in the B complex, resulting in the destablization of U1 and U4 snRNPs and the formation of the catalytic B* complex. The B* complex carries out the first transesterification reaction before it is converted to the C complex, which catalyzes the second transesterfication step. Upon completion of the splicing reaction, the spliceosome dissembles and the snRNPs are thought to be recycled into other spliceosomes (84).

The three major sequences which contribute to defining an intron are usually short and degenerate in mammalian cells with the canonical 5' and 3' splice site sequences being CAG/guragu (/ is the exon/intron boundary, R is A or G) and uuuyyyyyunyag/G (y is C or U, N is any nucleotide), respectively (85). This very limited amount of sequence information poses a

great challenge for the spliceosome, which must recognize intron/exon boundaries with great accuracy and fidelity. This conundrum is solved by the integration of an extensive second layer of splicing regulatory elements, including intronic splicing enhancers/silencers (ISE, ISS), exonic splicing enhancers/silencers (ESE, ESS) and a growing roster of *trans*-acting protein factors. Together, they modify the strength of individual splice sites, thus determining splice site selection and influencing both constitutive and alternative splice site usage (Figure 3-3).

Although the line between constitutive and alternative splice sites is not necessarily clear-cut, there are subtle qualitative differences between the two groups. In terms of the three primary sequence elements that define a splice site, the majority of alternative splice sites are suboptimal in that those sequences deviate more strongly from the consensus sequences compared to constitutive splice sites (82,86). Alternative 5' splice sites tend to deviate from the consensus CAG/guragu at the +4 and +5 positions, which would most likely impair efficient interaction with U1 and U6 snRNPs, while alternative 3' splice sites in general have more purine-rich polypyrimidine tracts which may reduce its affinity for U2AF65 (82,85-87). Interestingly, exons expressed exclusively in neurons and muscle cells, as well as exons that are developmentally regulated, show even stronger deviation from the consensus compared to other classes of alternative exons (82,86).

The suboptimal characteristics of alternative splice sites open a window for splicing regulation. Indeed, various intronic and exonic elements act to promote or inhibit individual splice site utilization through their interactions with protein factors. The best characterized ESEs are purine-rich sequences which facilitate spliceosome assembly by binding to SR proteins, and the best studied ISSs and ESSs are the ones bound by heterogeneous ribonucleoproteins (hnRNPs) (88), such as hnRNP A1 and hnRNP I (also known as polypyrimidine tract binding

protein, PTB) (89-92). PTB can bind to pyrimidine rich sequences at the 3' splice sites of some alternatively spliced exons, including c-src, α-actinin and human α-tropomyosin pre-mRNAs, and repress splice site utilization by blocking U2AF recruitment (93). Yet, the simple view of defining SR proteins as splicing enhancers and hnRNPs as splicing silencers failed to explain the regulation of many alternative exons (94-97). Most importantly, the same splicing factor can exert opposite functions in the context of different pre-mRNAs, in many cases the outcome depends on the position of the factor binding sites relative to the regulated exons (98).

A single exonic/intronic element together with its corresponding *trans*-acting factor(s) can regulate splicing autonomously. Yet, it is not uncommon that an alternative exon is regulated by multiple, often antagonistic, elements (99). The relative concentrations of these *trans*-acting factors could dictate the final splicing outcome. Therefore, tissue-specific or developmental-stage-specific splicing patterns might arise, at least in part, as a result of the unique splicing factor expression profile in the cell. A good example is the splicing reprogramming in neuronal differentiation induced by the PTB/neuronal PTB (nPTB) switch. PTB and nPTB are highly homologous proteins encoded by two paralogous genes. However, compared to PTB, nPTB is a significantly weaker splicing repressor for several neuron-specific exons (100-102). During neuronal differentiation, PTB protein levels decrease while nPTB increases, reprogramming splicing toward neuronal specific patterns (103,104).

The alternative splice site choice can be influenced not only by the expression levels of certain splicing factors, but also by their posttranslational modifications (105), thus opening up a whole new venue for cell signaling in the regulation of alternative splicing (106). For example, M-phrase splicing repression during the cell cycle is partly regulated by splicing factor SRp38. SRp38 is phosphorylated throughout the entire cell cycle except in the M-phase during which

SRp38 is dephosphorylated in its RS domain which allows tight binding to U1 70K protein. This interaction prevents U1 70K from binding to other SR proteins which are required to activate splicing (105).

In summary, alternative splicing regulation involves a complex regulatory network, which requires positive and negative splicing regulators working in conjunction with relatively weak splice sites and flanking intronic and exonic regulatory sequences.

Alternative Splicing and Human Diseases

Dysregulated alternative splicing is a major cause of human disease. Mutations in *cis*-elements as well as *trans*-acting factors lead to disease-associated splicing patterns. *Cis*-acting mutations were originally underestimated to account for ~15% of human diseases. This estimate only reflects mutations that disrupt known splice sites (107). Recent reassessment of the figure revealed that >50% of disease-causing mutations result in aberrant splicing (108). Based on a probabilistic model, it has been proposed that ~60% of human diseases could be caused by *cis*-acting mutations that disrupt splicing codes (109). Although not as prevalent, mutations affecting trans-acting splicing factors also affect the integrity and dynamics of the splicing machinery and result in aberrant splicing. A number of examples are listed in Table 3-1.

Myotonic Dystrophy, Aberrant Splicing and Muscleblind-like Proteins

Myotonic dystrophy (DM) provides an excellent example where mis-regulated alternative splicing causes inherited human disease. At the molecular level, at least 21 pre-mRNAs have been shown to undergo aberrant splicing in DM tissues (110). Although the physiological readouts for the majority of the misregulated splicing events remain to be elucidated, direct cause-effect connections have been drawn between several aberrant splicing events and characteristic clinical features of DM. Aberrant retention of exon 7a in *CLCN1* mRNA introduces a pre-mature stop codon (PTC) which targets the mRNA to the nonsense-mediated

decay (NMD) pathway (111,112). Reduction in *CLCN1* mRNA levels leads to loss of this major skeletal muscle chloride channel, accounting for the reduced Cl conductance and myotonia in DM patients. Another example of mis-splicing in DM is the insulin receptor (IR). Elevated IR exon 11 inclusion in DM skeletal muscle results in predominant expression of an insulin-insensitive isoform of IR which leads to the insulin insensitivity observed in DM patients (113,114).

Although the splicing patterns of many pre-mRNAs are abnormal in DM, there is no general affect on splicing efficiency or fidelity. In fact, DM selectively affects a group of exons that undergo developmental splicing switches as illustrated in Figure 3-4. Interestingly the aberrant splicing pattern in DM invariably correlates with the corresponding embryonic splicing pattern. Thus, the activity of a certain splicing factor which is responsible for maintaining normal adult splicing patterns may be missing in DM. Several studies have shown that CUG triplet repeat RNA binding protein 1 (CUGBP1) is able to regulate the alternative splicing of a number of misregulated exons in DM, including cTNT exon 5, IR exon 11 and CLCN1 exon 7a (112,113,115). However, the embryonic or DM-like splicing patterns are seen upon CUGBP1 over-, instead of under-expression, suggesting that DM is not caused by CUGBP1 loss-of-function.

In contrast, the MBNL1 sequestration model proposes that MBNL1 loss-of-function underlies DM pathogenesis, and is supported by both *in vitro* and *in vivo* evidence. Yet the normal function of MBNL1 was unknown when I began this project. As a working hypothesis, we proposed that MBNL1 functions to directly regulate developmental splicing switches and loss of MBNL function leads to the splicing defects in DM.

Results

Skeletal Muscle Fast Troponin T (Tnnt3) Undergoes a Postnatal Developmental Splicing Switch

Tnnt3 encodes a skeletal muscle specific troponin T. Alternative splicing of fast troponin T (Tnnt3) is abnormally regulated in DM patients as well as in *Mbnl1*$^{\Delta E3/\Delta E3}$ mice in that the fetal exon shows an elevated inclusion rate in DM patients (69). To test if alternative splicing of Tnnt3 is developmentally regulated, reverse transcription-polymerase chain reaction (RT-PCR) analysis was performed using RNAs isolated from the quadriceps of wild type (WT) mice aged from embryonic day 14.5 (E14.5) to postnatal day 28 (P28). PCR primers were positioned in exons 2 and 11. A number of exons represented as the black boxes in Figure 3-5 are developmentally regulated, and the alternative splicing of these exons contributed to multiple PCR products (Figure 3-5, uncut). It is interesting to note that the overall splicing of Tnnt3 undergoes a developmental transition postnatally in that multiple splice isoforms gradually consolidate into a single major isoform that predominate in P28 mouse muscle. To assay specifically for the splicing transition of the Tnnt3 fetal exon, PCR products were digested with BsrBI, which had a unique cleavage site inside the fetal exon. After BsrBI digestion, isoforms including/excluding the fetal exon could be distinguished according to their sizes. In P2 mice, the fetal exon was included in > 90% of Tnnt3 mRNAs whereas the fetal exon inclusion rate dropped nearly to zero in P28 mice, reflecting a narrow time window for this splicing transition. A high fetal exon inclusion rate in both neonates, as well as DM patients, is also consistent with the finding that DM patients adopt the fetal splicing pattern for a specific group of pre-mRNAs.

MBNL1 Promotes Tnnt3 Fetal Exon Exclusion

To test if MBNL1 can regulate the alternative splicing of Tnnt3 fetal exon, a Tnnt3 minigene was constructed by inserting the genomic region between mouse Tnnt3 exon 8 and

exon 9 into the vector pSG5 (Stratagene). The minigene reporter was transfected into 293T cells together with plasmids expressing green fluorescent protein (GFP) or GFP-tagged RNA-binding proteins. One to two days post-transfection, RNAs were isolated from the cells and RT-PCR was performed using primers flanking Tnnt3 E8 and E9. In the presence of GFP alone, the minigene showed a default <5% fetal exon exclusion. Overexpression of either GFP-MBNL1 full length or N-terminus, which contains the four CCCH motifs responsible for RNA binding (116), dramatically increased the fetal exon exclusion rate to >90% (Figure 3-6 A). This effect was specific to MBNL1, as equivalent amounts of GFP-CUG-BP1 or GFP-hnRNP A1 which often inhibits alternative exon inclusion(91,99,117), failed to significantly change the splicing pattern (Figure 3-6 A). Furthermore, the extent of fetal exon exclusion was MBNL1 dosage dependence. When increasing amounts of GFP-MBNL1 plasmids were cotransfected with the minigene, the percentage of fetal exon exclusion positively correlated with the level of GFP-MBNL1 proteins (Figure 3-6 B). Therefore, MBNL1 functionally promotes Tnnt3 fetal exon exlusion.

MBNL1 Physically Interacts with Tnnt3 Pre-mRNA

Although MBNL1 regulated Tnnt3 fetal exon splicing, it remained unclear whether the effect was primary or secondary. Since RNA-protein interactions are a pre-requisite for alternative splicing regulation, potential MBNL1-Tnnt3 pre-mRNA interactions were examined using a UV crosslinking approach. HEK293T cells were transfected with protein expression plasmids encoding myc-tagged versions of CUGBP1, full-length MBNL1 (MBNL1 FL) or the MBNL1 N-terminal region (MBNL1 N). Following transfection, cell lysates were incubated with radiolabelled Tnnt3 RNAs encompassing exons 7-9 or different 500-645 nucleotide (nt) subregions designated T1-6 (Figure 3-7 A), photocrosslinked with UV-light and digested with RNase A. Protein-RNA complexes were then immunopurified with the anti-myc monoclonal

antibody (mAb) 9E10 and resolved by SDS-PAGE. Because these studies indicated that only Tnnt3 T5 RNA (500 nt) bound to both MBNL1 FL and MBNL1 N proteins, this region was further subdivided into T5.1-T5.45. Interestingly, only T5.45 RNA (200 nt) crosslinked to CUGBP1 and MBNL1 proteins while T5.1 (125 nt), and the other subregions (T5.2 and T5.3, data not shown), did not (Figure 3-7 B). In agreement with prior studies, CUGBP1 failed to crosslink to a CUG^{exp}, $(CUG)_{54}$, while both MBNL1 FL and MBNL1 N did (68).

MBNL1 Binding Sites Are Required for MBNL1-Mediated Splicing Regulation

A recent mapping study identified several MBNL1 binding sites containing a core element within human and chicken cardiac troponin T (cTNT/TNNT2) (118). Alignment of these sequences revealed a hexanucleotide consensus motif (5'-YGCUU/GY-3', Y-C orU). A highly similar element (TGCGCTT) was found in Tnnt3 pre-mRNA immediately upstream of the fetal exon 3' splice site. In order to test whether an MBNL1 binding site was essential for MBNL1-mediated splicing regulation, a 10 nucleotide region encompassing the TGCGCTT element was deleted in the Tnnt3 minigene (Figure 3-8 A). Crosslinking assays were then used to test whether wild type and mutant Tnnt3/98 RNA (a 98-nt subregion encompassing 83-nt of the 3' end of intron 8 and 15-nt of the F exon) was recognized by both CUGBP1 and MBNL1 (Figure 3-8 B, lower panel). Concurrently, wild type or mutant Tnnt3 minigenes were transfected into C2C12 cells to assay whether Tnnt3 splicing remained responsive to MBNL1 overexpression (Figure 3-8 B, upper panel). These assays were performed using C2C12 myoblasts since they showed a higher default level of Tnnt3 F exon skipping compared to 293T cells. The resulting Δ10 mutant showed much reduced MBNL1 crosslinking compared to WT minigene (Figure 3-8 B, lower panel), suggesting at least one of the MBNL1 binding sites on Tnnt3 pre-mRNA was disrupted. When tested in the cotransfection assay, the Δ10 mutant showed a corresponding

impairment of F exon skipping promoted by MBNL1 overexpression (Figure 3-8 B, upper panel). Indeed, the Δ10 deletion eliminated F exon skipping in cells transfected with the Tnnt3 minigene alone or with CUGBP1-mycHis together with the minigene (Figure 3-8 B, RT-PCR) suggesting loss of MBNL1 binding and enhanced spliceosome recruitment to the F exon region. Although MBNL1 binding was not completely eliminated by deleting the 10 nucleotides, the reduced fetal exon splicing response upon MBNL1 overexpression suggests that the MBNL1 binding site was essential for MBNL1-mediated splicing regulation.

Conclusions, Discussion and Future Work

MBNL1 Protein Regulates Alternative Splicing

MBNL1 bound directly to an intronic sequence immediately upstream of the Tnnt3 fetal exon to promote fetal exon exclusion. Furthermore, an intact MBNL1 binding site was necessary for efficient enhancement of fetal exon exclusion promoted by MBNL1. Therefore, we concluded that MBNL1 functions as a potent alternative splicing factor that promotes the adult splicing patterns for specific target pre-mRNAs. This function of MBNL1 appears to be conserved from the *Drosophia* homolog, which regulates the splicing of α-actinin (119). Although only one pre-mRNA, Tnnt3, was shown in this chapter to be regulated by MBNL1, a number of other transcripts including cTNT, IR and sarcoplasmic reticulum Ca^{2+}-ATPase 1 (Serca1) have been shown to be direct MBNL1 targets (118,120). For all of these pre-mRNAs, MBNL1 invariantly promotes the adult splicing pattern, regardless of whether splicing involves the inclusion of an adult-specific exon, as is the case for IR and Serca1, or the exclusion of a fetal exon, as is the case for Tnnt3 and cTNT. Interestingly, the positions of MBNL1 binding sites relative to these regulated exons correlate with the splicing patterns promoted by MBNL1. For example, in the case of Serca1, where MBNL1 promotes adult exon inclusion, MBNL1

binding sites have been mapped to intronic sequences downstream of the alternative exon (120). On the other hand, for cTNT and Tnnt3 where MBNL1 promotes exon exclusion, MBNL1 binds to intronic sequences immediately upstream of the respective fetal exons. A parallel scenario was found in the case of another alternative splicing factor, neuro-oncological ventral antigen 1 (Nova1), where the splicing patterns promoted by Nova1 could be accurately predicted based on the positions of Nova binding sites (98,121). As with MBNL1, Nova1 usually promotes exon exclusion when binding immediately upstream of the regulated exon while it promotes exon inclusion when binding downstream of the alternative exon. Although the mechanisms by which these two proteins regulate alternative splicing remain elusive, there could be significant similarities between them. For instance, both proteins may promote exon exclusion by binding upstream of the 3' splice site and blocking U2AF recruitment and subsequent spliceosome assembly. Further studies are required to elucidate the mechanism by which MBNL1 promotes exon inclusion/exclusion.

MBNL2 and MBNL3 have been shown to promote adult splicing patterns to the same extent as MBNL1 in the cases of cTNT, IR and Tnnt3 minigenes ((118) and data not shown). While MBNL3 expression is mainly confined to placenta among adult human tissues, MBNL2 expression, as assayed by Northern blotting, is higher than MBNL1 across a variety of tissues both in mouse and in human (79,80). Yet intriguingly, *Mbnl2* knockout mice failed to recapitulate major DM symptoms (122). The reason for this failure is unknown. Due to the lack of a good MBNL2 antibody, the levels of MBNL2 protein in tissues remain to be assessed. One possibility is that MBNL2 translation is repressed although it is being actively transcribed, so that the major MBNL protein across tissues is MBNL1. A similar situation has been described for PTB and nPTB, where both genes are actively transcribed in neuronal precursor cells, yet

nPTB is under translational repression by micro RNAs (miRNAs), so that only PTB protein predominates in undifferentiated neuronal precursors and represses neuronal-specific splicing patterns (104,123).

The Role of MBNL1 in DM Pathogenesis

The MBNL sequestration model proposes that the expression of expanded CUG or CCUG repeats in DM are toxic because they titrate cellular MBNL proteins, leading to their loss-of-function and subsequent DM phenotypes. Although several lines of evidence support the hypothesis that pathogenic RNAs functionally sequester MBNL, there was a missing link between MBNL loss-of-function and DM phenotypes. In this study, I have demonstrated that MBNL1 acts as an alternative splicing factor and promotes the adult splicing pattern of DM-affected transcripts. As illustrated in Figure 3-9, pathogenic RNAs carrying expanded CUG or CCUG repeats titrate MBNL, which leads to reduced levels of available MBNL proteins in the nucleoplasm. Thus, antagonistic splicing factors, such as CUGBP1, can more efficiently promote the fetal or DM-like splicing patterns of certain transcripts in adult tissues. The resulting ectopic expression of fetal isoforms in various adult tissues and organs eventually leads to the multisystemic symptoms observed for DM patients.

The Importance of Developmental Splicing Regulation

Although extensive work has been done on developmentally regulated gene transcription, the developmental regulation of splicing is an area that has been overlooked in the past. More evidence has emerged during the last decade, which shows that a number of transcripts undergo splicing switches at a certain stage during development. Differences between fetal and adult splicing patterns, as in the case of *CLCN1*, have been shown to affect mRNA stability, leading to drastically different protein levels. Other mis-splicing events, as in the case of *IR*, have been shown to generate functionally distinct protein isoforms that are specifically compatible with

either fetal or adult tissue functions (111-113). However, the direct effects of the majority of the developmental regulated splicing events remains to be explored. Even for those splicing events with known functional consequences, little is know about their biological relevance. For example, why do neonatal muscles require fewer chloride channels to function properly? Are there major differences in the molecular environment between fetal and adult tissues so that only fetal/adult isoforms would fit in the fetal/adult environment? How conserved are these events evolutionarily? At what time during development do specific splicing switches occur? Do all developmental splicing switches happen during the same time window or are there multiple sets of such changes taking place during developmental progression? Are MBNL/CELF proteins the master regulators of all such switches? If not, what are the other pathways? Systematic and high-throughput approaches need to be utilized to identify developmentally regulated splicing events across tissues and organs in various model organisms. Functional grouping of these events could potentially shed light on their biological significance. If, hypothetically, the majority of these events involve proteins functioning in muscle contraction and neuronal signaling, the functional significance of these splicing switches might lie mainly in assisting animals to adapt to the distinct demands of fetal versus adult lives associated with changing mobility and evolving tasks of learning and memory. Combined with human genetic analysis, single nucleotide polymorphisms (SNPs) at these developmentally regulated splice sites could be assayed for their correlation with human diseases.

Future Work

Although it is now clear that MBNL1 functions as an alternative splicing factor, the normal splicing targets of MBNL1, as well as the mechanism by which it promotes exon inclusion/exclusion, remains unclear. Using an *in vivo* cross-linking and immunoprecipitation technique (124), a greater number of MBNL1 targets, as well as MBNL1 binding sites on those

targets, can be identified in a more comprehensive manner. It would be interesting to test whether those newly identified pre-mRNAs also undergo developmental splicing switches, and whether these switches are MBNL1-dependent. To address the mechanistic question, *in vitro* splicing assays of MBNL1 targets could be performed in the presence or absence of MBNL1 so that subsequent formation of spliceosomal intermediate complexes could be compared which may give clues to the mechanism of MBNL1 mediated splicing regulation. In the cases where MBNL1 promotes exon inclusion, additional factors might be recruited by MBNL1 to enhance spliceosome assembly. Immunoprecipitation could be carried out to identify protein factors associated with MBNL1 in the cell which may mediate MBNL1-enhanced splicing events.

Table 3-1. Splicing factors and human diseases.

Disease	Affected splicing factor	Comments
Neuromuscular Diseases		
Myotonic dystrophy (DM)	Muscleblind-like protein 1, 2 and 3 (MBNL1, MBNL2, MBNL3)	Sequestration of human muscleblind proteins in DM foci disrupts MBNL mediated developmental stage-specific alternative splicing (68,69).
Myotonic dystrophy (DM)	CUG triplet repeat RNA-binding protein 1 and 2 (CUG-BP1, CUG-BP2)	CUG-BP1 and CUG-BP2 regulate alternative splicing in ways antagonistic to muscleblind proteins; elevated CUG-BP1 protein level has been reported in DM1 patients (112,115,125-127).
Spinal muscular atrophy (SMA)	Survival of motor neuron 1 and 2 (SMN1, SMN2)	Deletion/mutations in SMN1 and SMN2 genes results in insufficient level of SMN, leading to abnormal snRNP biogenesis (128,129).
Spinal cerebellum atrophy 2 (SCA2)	Ataxin-2	Ataxin-2 is a RNA-binding protein that interacts with neuronal splicing factor A2BP1 (Fox-1) (130).
Spinal cerebellum atrophy 8 (SCA8)	MBNL?	Drosophila muscleblind was identified as one of the poteintial modifier of a fly SCA8 model (131).
Spinal cerebellum atrophy 10 (SCA10)	Polypyrimidine tract binding protiein (PTB)?	AUUCU repeat RNA may sequester PTB; potential splicing defects in patients have not been tested (132).
Huntington's disease-like 2 (HDL2)	MBNL	MBNL1 shows decreased nucleoplasmic level while is enriched in CUG repeats containing nuclear foci (133)..
Fragile X tremeor ataxia syndrome (FXTAS)	Heterogeneous nuclear ribonucleoprotein A2 (hnRNP A2), MBNL1	hnRNP-A2 and MBNL1 are both enriched in the CGG repeat RNA containing nuclear inclusions (134).
Fascioscapulohumoral dystrophy (FSHD)	FRG1	Deletion of FRG1 transcriptional silencer leads to FSHD; mouse model overexpressing FRG1 phenocopy FSHD (135).
Paraneoplastic opsoclonus-myoclonus-ataxia (POMA)	Neurooncologic ventral antigen 1 and 2 (Nova-1 and Nova-2))	Nova is inactivated by autoantibodies; Nova knockout mice phenocopy the human POMA (136,137).
Other Neurological Disorders		
Isolated case of mental retardation and epilepsy	Ataxin-2 binding protein (A2BP1 (FOX-1))	A chromosome translocation disrupted A2BP1. Splicing patterns have not been tested in patients (138).
Retinitis pigmentosa (RP)	Precursor mRNA-processing factor 3, 8, 31 (PRPF3, PRPF8, PRPF31)	Loss of constitutive splicing factors PRPF3, PRPF8, PRPF31 leads to RP (139-141)..
Cancer		
Leukemias and sarcomas	Translocation liposarcoma (TLS)	Fusion of TLS with ETS-related gene (ERG) or FUS eliminates its interaction with SR proteins, inhibiting SR protein mediated CD44 splicing (142,143).
Azospermia	RNA binding motif protein, Y-linked, family 1, member A1,	RBMY interacts with Tra2 and regulates alternative splicing (144).
Hepatocellular carcinoma	HCC1	Autoantibodies against HCC1 were detected in hepatocelluar carcinoma patients (145).

Table 3-1. Continued.

Disease	Affected splicing factor	Comments
Papilary renal cell carcinoma	Splicing factor, proline- and glutamine- rich (SFPQ)	Fusion of SFPQ and $p54^{nrb}$ to TFE3 gene was identified in papillary renal cell carcinoma lines (146).
Developmental Disorders		
Rett syndrome (RTT)	methylation-dependent transcriptional repressor methyl-CpG binding protein 2 (MeCP2)	MeCP2 regulates alternative splicing through interaction with RNA-binding protein YB-1 (147).
Prader Willi Syndrome (PWS)	HBII-52	Loss of HBII-52 correlates with aberrant serotonin receptor splicing in PWS patients (148).
Hay-Wells syndrome	p63 alpha	p63 alpha mediates FGFR-2 alternative splicing through interaction with ABBP1 (149).

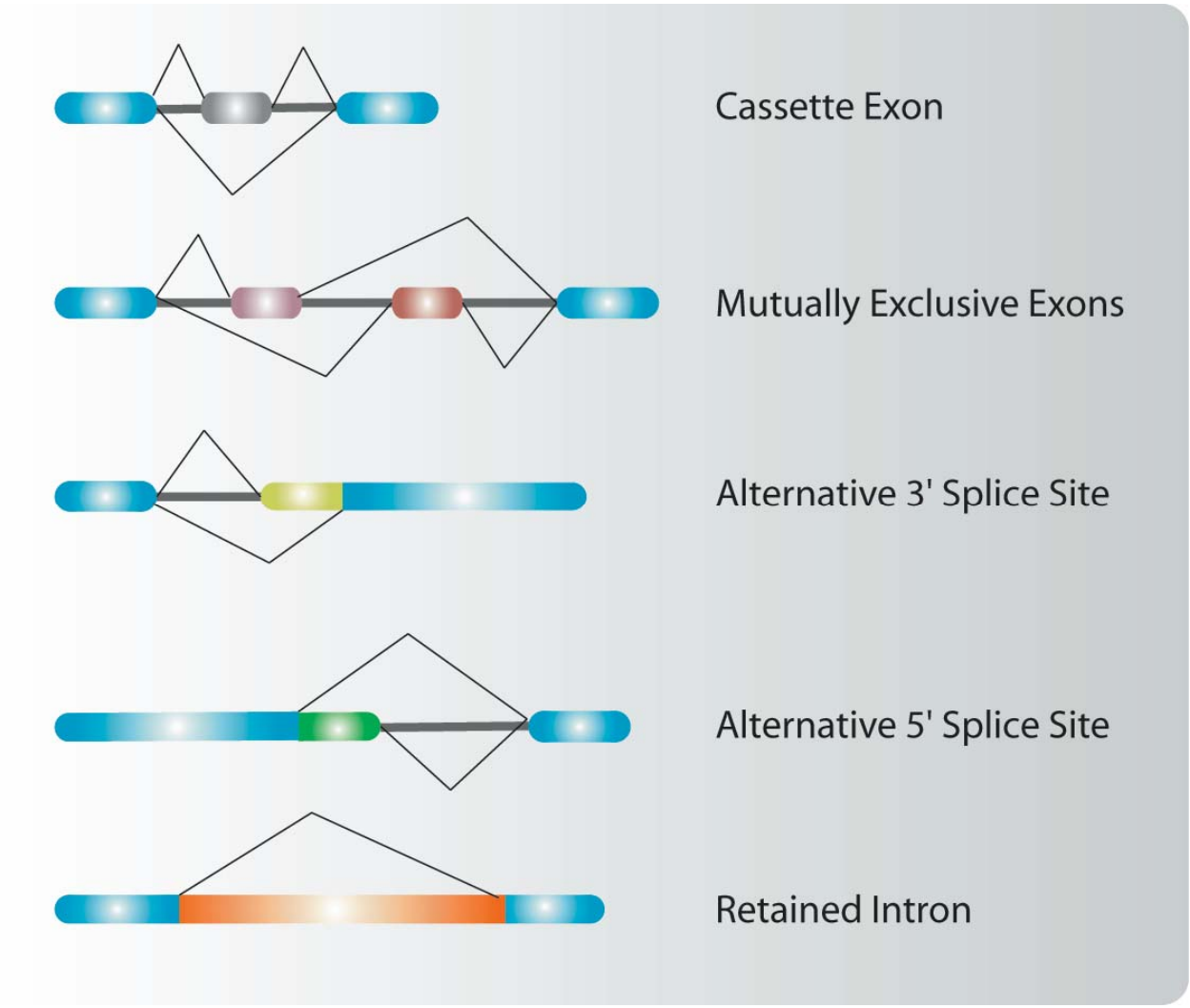

Figure 3-1 Types of alternative splicing events. Constitutive exons are shown in blue, alternative exons are shown in grey, purple, red, yellow, green and orange. Introns are shown as grey horizontal lines connecting exons. Thin grey lines indicate splicing events.

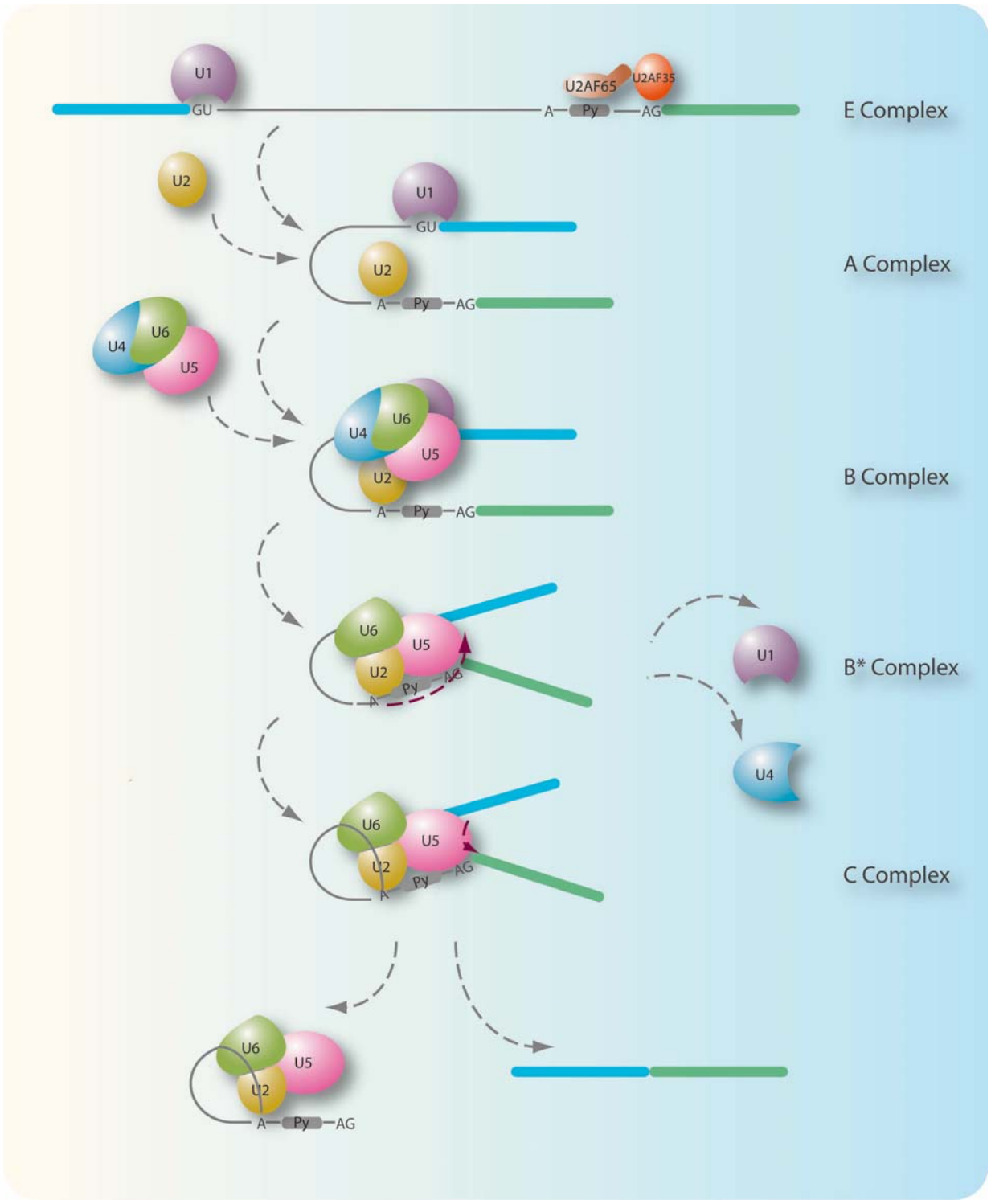
U1
GU
U2AF65
U2AF35
A
Py
AG
E Complex
U2
U1
GU
U2
A
Py
AG
A Complex
U4
U6
U5
U4
U6
U5
U2
A
Py
AG
B Complex
U6
U5
U2
Py
AG
A
U1
U4
B* Complex
U6
U5
U2
Py
AG
A
C Complex
U6
U5
U2
A
Py
AG

Figure 3-2 Splicing Mechanism. U1 snRNP is first recruited to the 5' splice site through base pairing with the 5' splice site, forming the spliceosomal E complex. With the help of U2AF, U2 snRNP stably associates with the branch point sequence (BPS), which forms the A complex. Subsequent recruitment of U4/U5/U6 tri-snRNP generates the pre-catalytic B complex. Major rearrangements of RNA-RNA and RNA-protein interactions occur in the B complex, resulting in the destablization of U1 and U4 snRNPs and the formation of the catalytic B* complex. The B* complex carries out the first transesterification reaction before it is converted to C complex, which catalyzes the second transesterfication step. Upon completion of the splicing reaction, the spliceosome dissembles and the snRNPs are thought to be recycled into other spliceosomes.

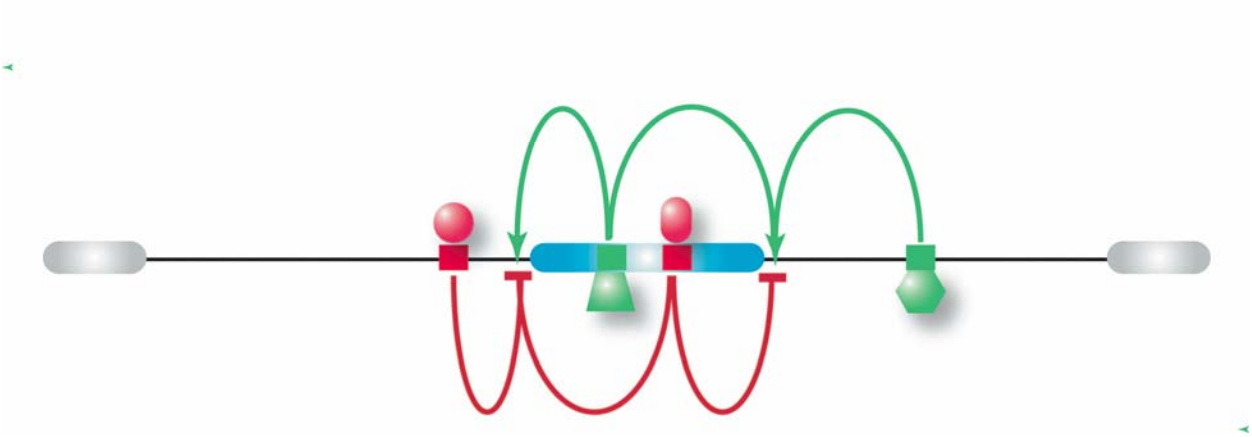

Figure 3-3 Regulation of alternative splicing. Constitutive exons are shown as grey bars. The alternative exon is shown in blue. Introns are shown as black lines connecting the exons. Intronic/exonic splicing silencers (ISS, ESS) and corresponding *trans*-acting factors are shown in red. Intronic/exonic splicing enhancers (ISE, ESE) and interacting factors are shown in green.

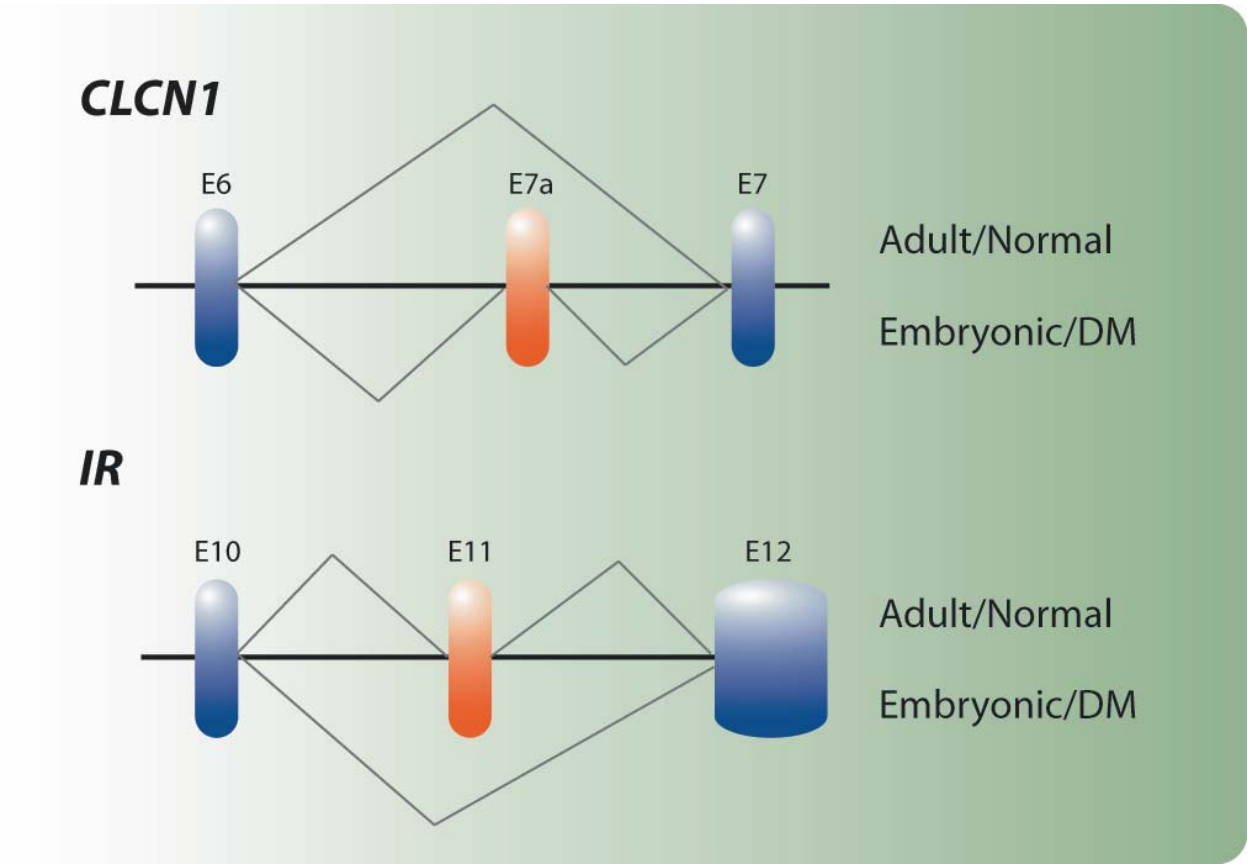

Figure 3-4 Illustration of aberrant splicing patterns in DM. Constitutive exons are shown in blue and alternative exons are shown in orange. Introns are shown as black horizontal lines connecting the exons. Black thin lines indicate splicing events.

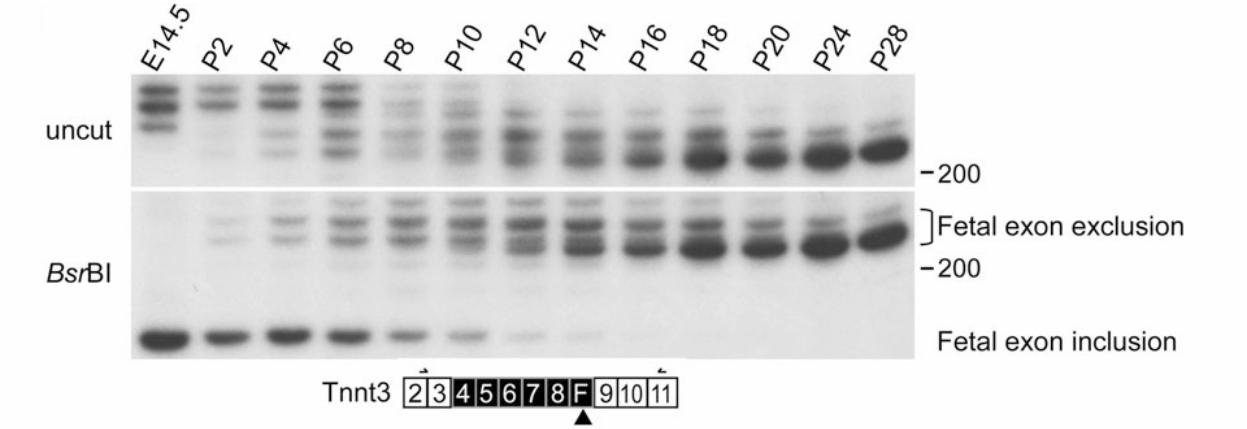

Figure 3-5 Tnnt3 undergoes a postnatal developmental splicing switch. In WT mice, RT–PCR analysis of fast skeletal muscle troponin T (Tnnt3) mRNA showed alternative splicing of the fetal (F) exon as well as exons 4 through 8. Primers used for PCR are indicated as black arrows in Tnnt3 exon 2 and 11. Transition from inclusion to exclusion of the fetal exon occurred mainly between P2 and P16. The lower panel shows RT–PCR products after digestion with BsrBI, which cleaves within the fetal exon. E14.5 is whole limb while P2-P28 are dissected quadriceps.

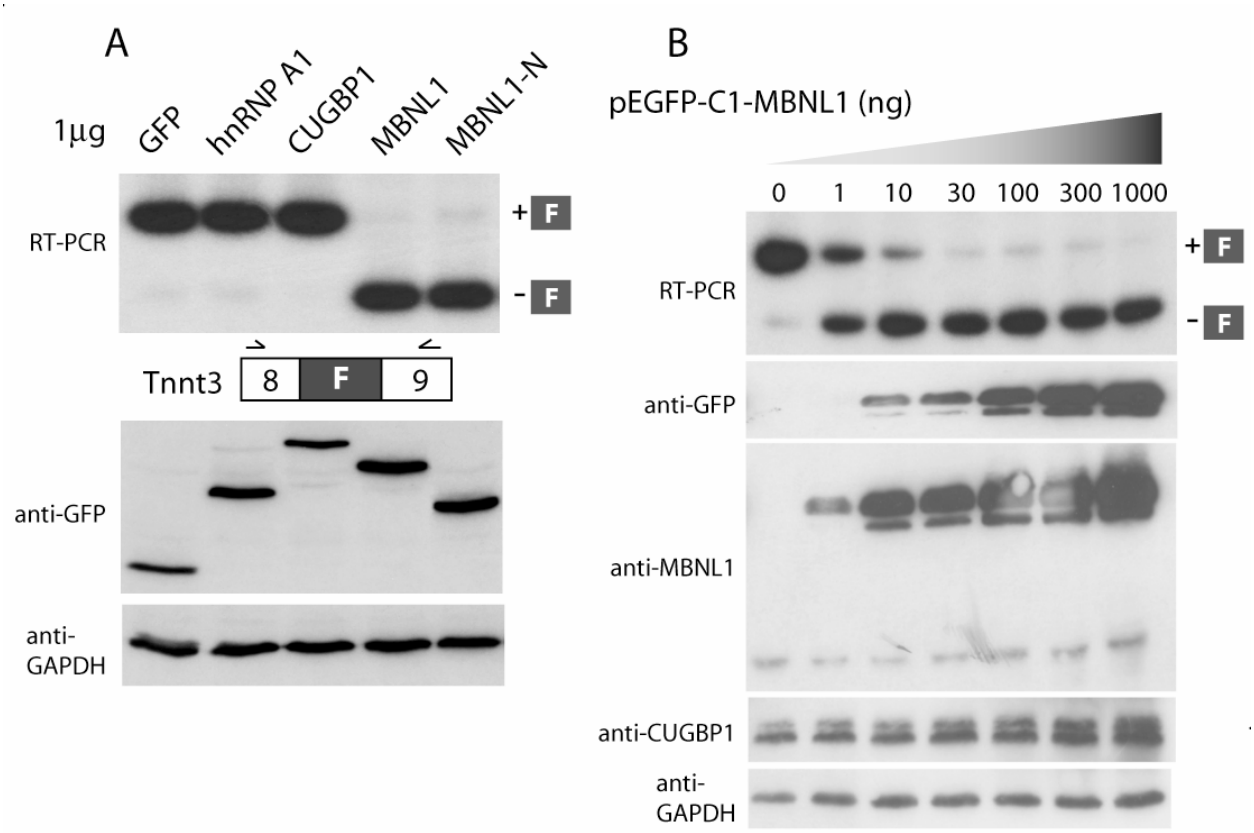

Figure 3-6 MBNL1 promotes Tnnt3 fetal exon exclusion. The Tnnt3 minigene was cotransfected into HEK293T cells with plasmids expressing GFP or GFP-tagged RNA-binding proteins. RT-PCR shows the ratios of Tnnt3 fetal exon inclusion/exclusion. Expression levels of GFP fusion proteins are shown by immunoblotting with anti-GFP and anti-MBNL1 antibodies. GAPDH is the protein loading control. **A**) GFP-MBNL1 promotes Tnnt3 fetal exon exlusion. Overexpression of GFP-MBNL1 full length as well as the N-terminus increased fetal exon exclusion rate compared to overexpression of GFP alone. GFP-hnRNP A1 or GFP-CUGBP1 do not promote fetal exon exclusion while expressed at comparable levels with GFP-MBNL1. **B**) Tnnt3 fetal exon exclusion is MBNL1 dosage-dependent. Increasing amounts of GFP-MBNL1 were expressed in HEK 293T cells. Tnnt3 fetal exon exclusion rates display a corresponding increase with cellular MBNL1 levels. Overexpression of MBNL1 does not decrease CUGBP1 levels.

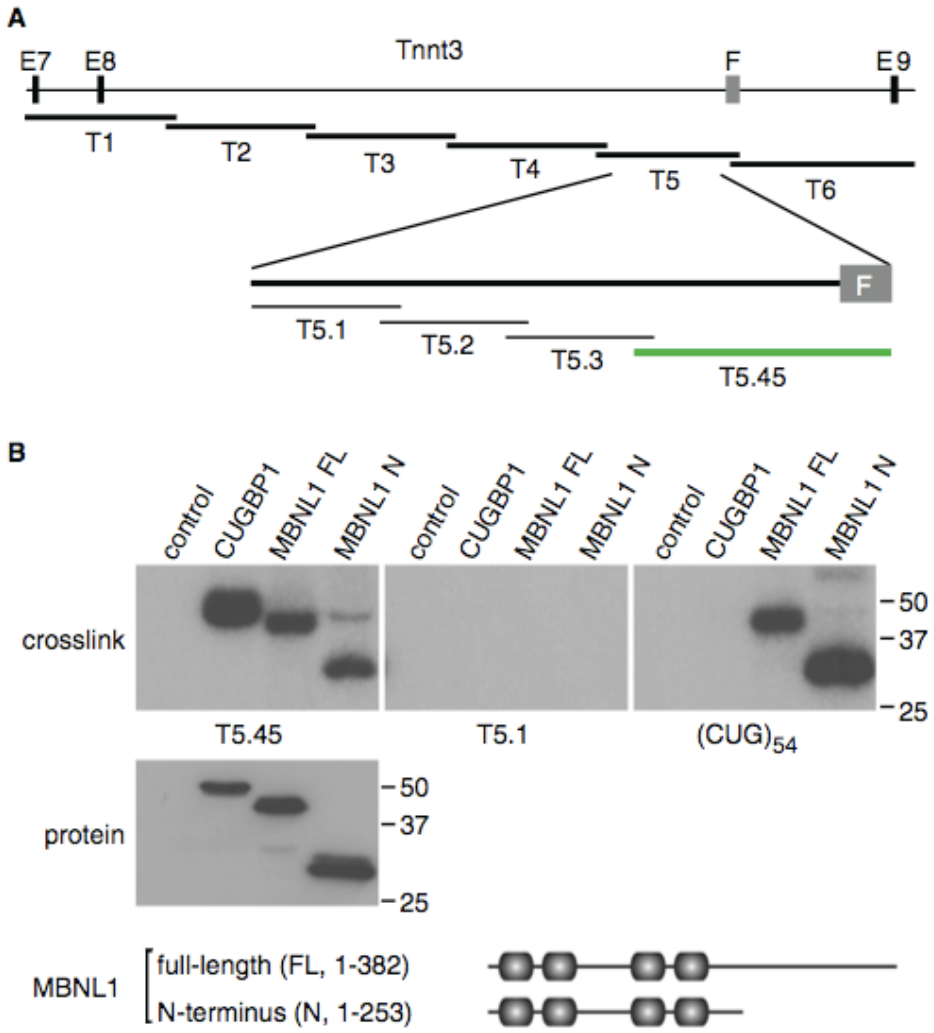

Figure 3-7 MBNL1 crosslinks to Tnnt3 pre-mRNA. **A**) Mapping of MBNL1 binding sites on Tnnt3. Illustration shows the Tnnt3 genomic region between exons 7 (E7) and 9 (E9) (thin lines, introns; black boxes, exons) and the subregions corresponding to the RNAs tested for MBNL1 crosslinking (thick lines, T1-6; thin lines, T5.1-T5.3; green line, T5.45). **B**) Interaction sites between MBNL1 and CUGBP1 on Tnnt3 pre-mRNA are positioned in a region which includes the F exon and upstream intron 8. HEK293T cells were transfected with myc-His tagged CUGBP1, MBNL1 FL or MBNL1 N followed by incubation of the corresponding cell lysates with ^{32}P-labeled Tnnt3 T5.45, T5.1 or $(CUG)_{54}$ RNAs and subsequent UV-light induced crosslinking and RNase A treatment. Labeled proteins were detected by SDS-PAGE and autoradiography (upper panels, crosslink) while protein levels in the lysates were detected by immunoblotting with the anti-myc mAb 9E10 (lower panel, protein). The primary structures of MBNL1 FL and MBNL1 N are illustrated below the immunoblot with the four CCCH motifs highlighted (shaded rectangles).

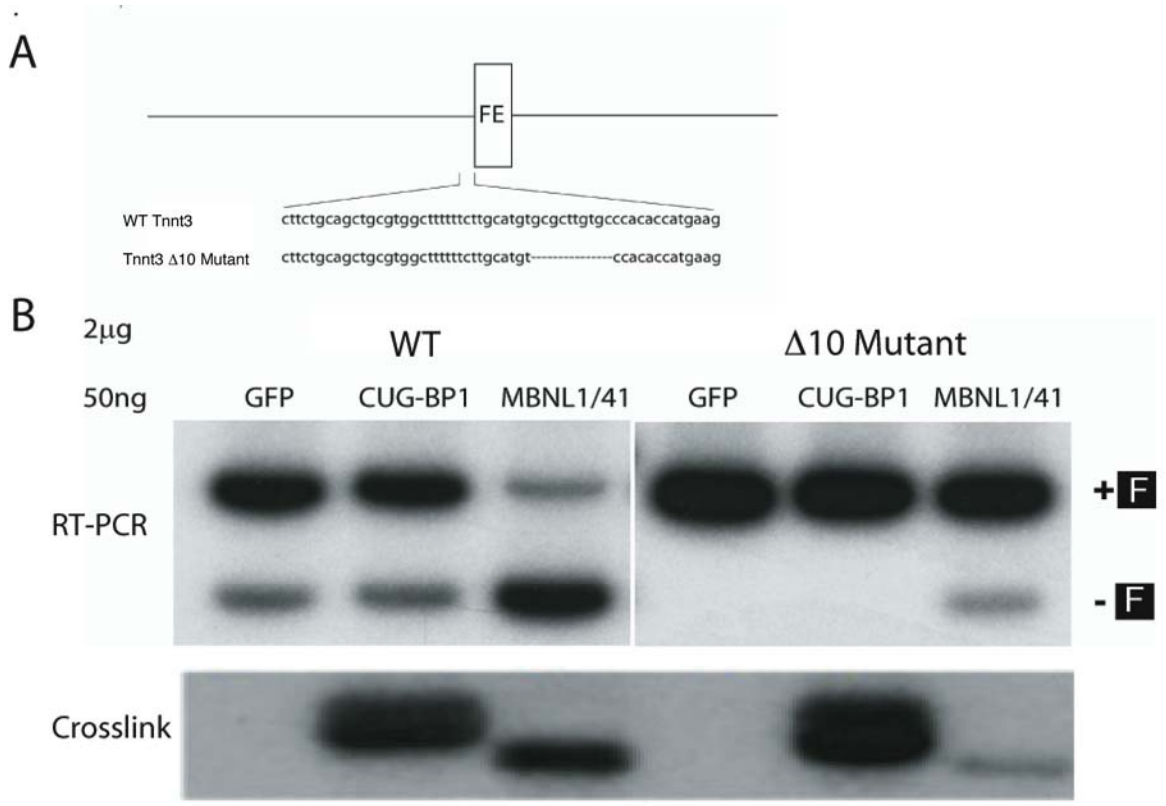

Figure 3-8 MBNL1 binding sites are required for MBNL1-mediated splicing regulation. **A**) Schematic views of the 10 nt deletion. **B**) The Δ10 mutant shows reduced MBNL1 crosslinking as well as MBNL1-mediated fetal exon exclusion. C2C12 myoblasts were used in the cotransfection assay since they show a higher default level of fetal exon exclusion compared to 293T cells.

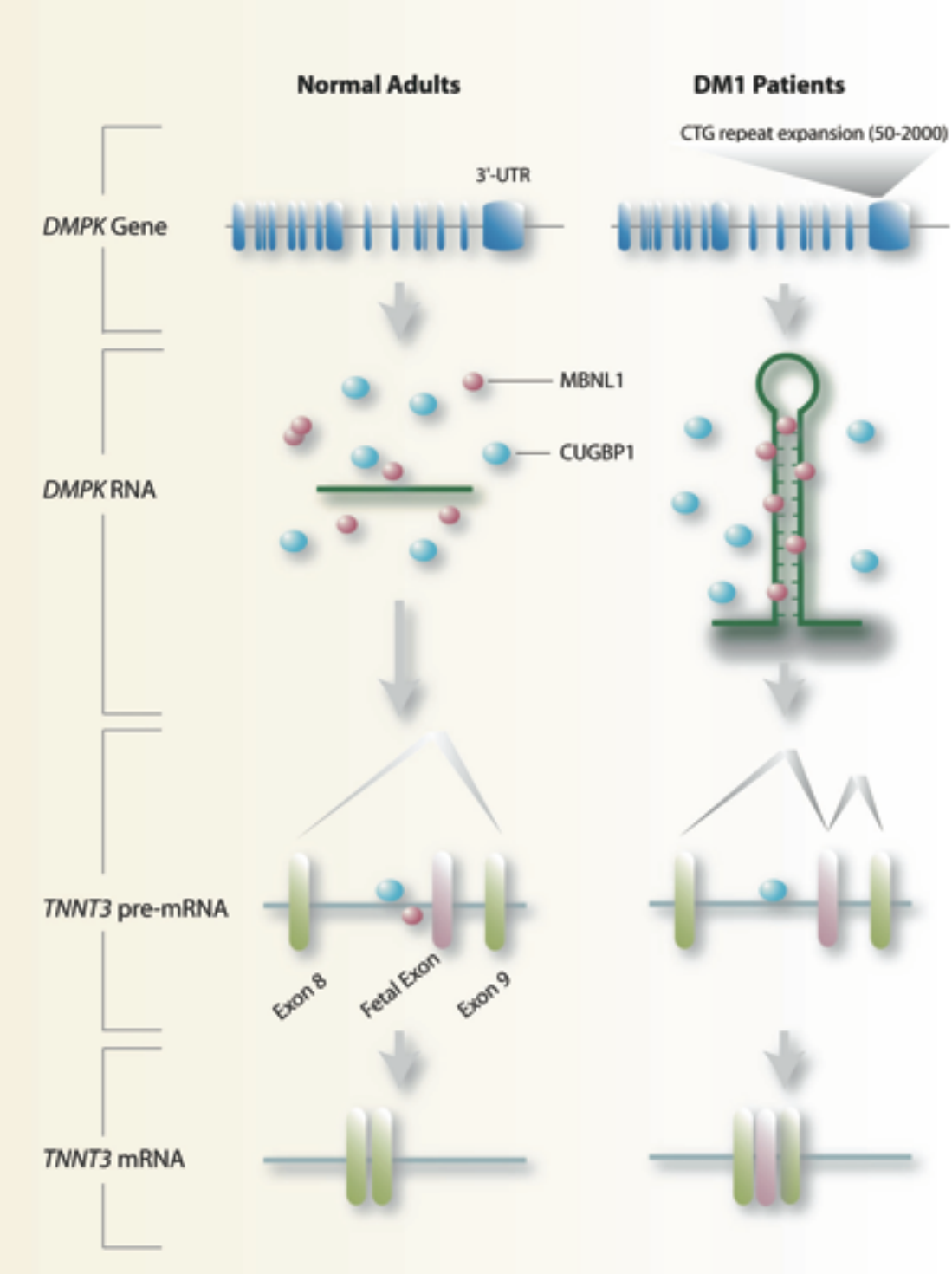

Figure 3-9 DM pathogenesis model. Pathogenic RNAs form hairpins and titrate MBNL proteins in DM cells. With reduced levels of available MBNL, antagonistic factors, such as CUGBP1, can more efficiently promote embryonic splicing patterns for certain pre-mRNAs in adult tissues, which leads to the DM phenotype.

CHAPTER 4
POTENTIAL MECHANISM FOR FUNCTIONAL SEQUESTRATION OF MBNL1 IN DM

Introduction

In the last chapter, evidence was presented supporting the idea that MBNL1 functions as an alternative splicing factor that regulates developmental splicing switches. A model was suggested in Figure 3-8 in which pathogenic RNA repeats sequester MBNL1 protein, leading to its loss-of-function and the downstream splicing defects. However, one question remains: what is the mechanism of MBNL1 functional sequestration in DM? In other words, what is the difference between the interactions between MBNL1 and pathogenic RNAs versus splicing targets? Several possibilities exist. First, MBNL1 might be functionally sequestered by CUG/CCUG repeats due to higher MBNL1 binding capacities on these pathogenic RNAs compared to those on MBNL1 normal splicing targets. However, this speculation is not compatible with the study by Ho *et al*. where 960 CUG or CAG repeats were transfected into COSM6 cells and assayed for their abilities to disrupt splicing {Ho, 2005 #11}. While MBNL1 relocalizes to both CUG and CAG RNA foci, only $(CUG)_{960}$ disrupted MBNL1-mediated splicing. The observation that CAG RNA repeats, which are of the same length and expression level as CUG repeats and thus might harbor comparable numbers of MBNL1 binding sites, failed to functionally titrate MBNL1 protein suggests that the expanded length of pathogenic RNAs is not the key reason underlying MBNL1 functional sequestration.

Two hypotheses remain, which are both based on the potential qualitative differences of MBNL1 interactions with pathogenic RNAs versus normal splicing targets. One proposes that MBNL1 is functionally sequestered due to its higher affinities for pathogenic RNA repeats compared to normal splicing targets. The other hypothesis proposes the distinct stabilities of ribonucleoprotein (RNP) complexes formed between MBNL1 and pathogenic RNAs versus

normal splicing targets. However, little was known about MBNL1-RNA interactions. Previous studies have suggested that the CCCH motifs ($CX_7CX_{4-6}CX_3H$) in MBNL proteins are related to those found in tris-tetrapoplin (TTP)/TIS11, a factor that promotes the deadenylation and turnover of target mRNAs containing a 3' untranslated region (3' UTR) class II AU-rich element (ARE) {Blackshear, 2002 #415}. Similar to TTP/TIS11d, the CCCH motifs in MBNL proteins are organized as tandem pairs separated by a 14-16-residue linker {Miller, 2000 #39; Nykamp, 2004 #420; Pascual, 2006 #225}. While TTP/TIS11d recognize single-stranded AU rich RNA sequences, MBNL1 has been shown to bind to double-stranded pathogenic RNAs as well as predicted single-stranded regions in cTNT pre-mRNA {Ho, 2004 #10}. However, the affinity of MBNL1 for different RNAs and the molecular details of MBNL1-RNA complexes remained to be characterized. In this chapter, a combination of chemical and biochemical methods were used to examine various aspects of MBNL1-RNA interactions. These results support the proposed mechanism of MBNL1 functional sequestration.

Results

MBNL1 Interacts with an Intronic Stem-Loop Structure Upstream of the Tnnt3 Fetal Exon

As mentioned in the previous chapter, identification of MBNL1 binding sites on cTNT revealed a hexanucleotide consensus motif (5'-YGCUU/GY-3'), which was located in a predicted single-stranded (ss) RNA regions. Together with prior observations that MBNL proteins are sequestered by CUG^{exp} and $CCUG^{exp}$ hairpins in ribonuclear foci in DM cells {Mankodi, 2003 #131}{Mankodi, 2005 #124}{Mankodi, 2001 #38}{Miller, 2000 #39}, this finding suggests that the MBNL proteins bind to both ssRNA and dsRNA structural motifs.

To determine if MBNL1 recognizes primarily ssRNA targets in other splicing precursors, the MBNL1 binding site on fast skeletal muscle troponin T (Tnnt3) RNA was mapped in greater detail. First, the structure of a 151-nt subregion of T5.45 was chemically and enzymatically

probed. The Mfold {Zuker, 2003 #315} predicted structure of this RNA fragment is shown in Figure 4-1 A, right panel. A set of three well-characterized chemical and enzymatic probes was used for RNA structure probing. Lead ion (Pb) induces RNA cleavage at ssRNA as well as relaxed dsRNA regions. Ribonuclease T1 cleaves G-nucleotides present in ssRNA regions while S1 nuclease also cleaves ssRNA regions but without base specificity. Two different concentrations of each probe were used in order to discriminate cleavages site with different susceptibilities. Based on the lead ion and nuclease sensitivity patterns, the Mfold predicted structure was confirmed. Interestingly, the 10-nt region which was important for MBNL1-mediated fetal exon exclusion as discussed in the previous chapter, is located in an 18-nt stemloop region, as indicated by the strong S1 nuclease and RNase T1 cleavage signals at G94 and reduced nuclease sensitivities in the surrounding regions ranging from G82 to G107 (Figure 4-1 A). Deletion of the 10-nt region is predicted to eliminate the hairpin structure {Zuker, 2003 #315}. The crosslinking and splicing results obtained with the Δ10 mutant indicated that the 18-nt hairpin was a binding site for MBNL1 and that MBNL1-hairpin interactions might promote fetal exon skipping. Because both sequence and structural elements could contribute to efficient MBNL1 binding, a double C→G and U→G mutant was generated that eliminated the C-C and mismatch in the 18-nt hairpin, which also increases the stability of this stem-loop, and furthermore substituted a G for a U in the loop. These mutations are not predicted to alter the overall folding pattern of Tnnt3/98. This double point (gg) mutant showed considerably reduced MBNL1 crosslinking compared to wild type Tnnt3 (Figure 4-1B) confirming that this region was an MBNL1 binding site. Interestingly, a similar decrease in MBNL1 crosslinking activity was also observed for the single C→G stem substitution mutant (data not shown). While loss of MBNL1 binding should promote fetal exon splicing, inclusion activity was completely

eliminated (Figure 4-1C, right panel). Loss of fetal exon splicing activity could result from enhanced hairpin stability, or an increase in the purine content of the polypyrimidine (Py) region upstream of the fetal exon 3' splice site, and subsequent impairment of spliceosome asembly. Interestingly, another double mutant (G→A and C→U), which was predicted to preserved the 18-nt stem-loop structure while reducing the GC content in the stem, also showed reduced MBNL1 cross-linking compared to wild type (Figure 4-1B, au mutant) while MBNL1-induced fetal exon skipping activity was reduced similar to the Δ10 mutant (Figure 4-1C, au mutant). Overall, these studies suggest that MBNL1 prefers to bind to a GC-rich stem-loop containing a pyrimidine mismatch(es) in a normal splicing target.

MBNL1 Binds to the Stem Region of a Pathogenic dsRNA

Muscleblind-like proteins were originally characterized as nuclear factors which are recruited by CUG^{exp} RNAs {Miller, 2000 #39}. The predicted double-stranded nature of these pathogenic RNAs has been validated by chemical and nuclease mapping, thermal denaturation and electron microscopy {Napierala, 1997 #479; Sobczak, 2003 #112; Michalowski, 1999 #111}. While the binding of MBNL proteins to CUG^{exp} RNA is proportional to the predicted stem length, there is currently no direct experimental evidence that MBNL binds directly to the CUG^{exp} stem or that the RNA structure remains in the hairpin configuration following MBNL binding. Therefore, we initially used chemical and enzymatic structure probing of labeled RNAs to identify MBNL1 binding sites on $(CUG)_{54}$ RNA. RNAs were 5' end-labeled, subjected to either lead ion (Pb)-induced hydrolysis or RNase T1 digestion in the absence or presence of recombinant MBNL1, and the products were fractionated on denaturing polyacrylamide gels. As shown previously, lead ions cleave both ssRNA and relaxed dsRNA structures which yielded a uniform ladder that increased in intensity with increasing lead concentration (Figure 4-2, left

panel). As anticipated, addition of MBNL1 inhibited strand cleavage in a concentration-dependent manner and densitometry analysis failed to show significant regional differences in the cleavage pattern by MBNL1 suggesting uniform binding of this protein throughout the stem region. In contrast to lead, RNase T1 prefers to cleave after G nucleotides in single-stranded regions. Thus, incubation with RNase T1 resulted in strong cleavage at the terminal loop (Figure 4-2, right panel, G26-G29). Interestingly, terminal loop cleavage was unaffected by MBNL1 addition while stem cleavage was uniformly inhibited. Therefore, we concluded that MBNL1 interacts primarily with the stem region of CUG^{exp} RNAs.

Similar Affinities of MBNL1 for Splicing Precursor And Pathogenic RNAs

According to the RNA sequestration model, pathogenic RNAs out compete normal RNA binding targets for MBNL1 leading to loss of MBNL-mediated regulation of alternative splicing during postnatal development. However, it is not clear why MBNL1 accumulates on DM1 and DM2 expansion RNAs in ribonuclear foci. The most straightforward explanation is that MBNL1 has a higher affinity for DM pathogenic RNAs compared to its physiological RNA splicing targets. Since the crosslinking of MBNL1 to the Tnnt3 fetal exon 3' splice site region and CUG^{exp} RNAs appeared to be comparable (Figure 3-7B), we determined the relative affinities of MBNL1 for $(CUG)_{54}$, $(CAG)_{54}$ and Tnnt3/5.45 using recombinant MBNL1 protein in filter binding and gel shift assays.

The MBNL1 proteins used for this study were either MBNL1 FL or the C-terminal truncation mutant MBNL1 N. For the MBNL1 FL preparation, ~60% of the purified protein was full-length while the MBNL1 N protein preparation was homogeneous as determined by Coomassie blue staining and immunoblot analysis (Figure 4-3A). Because MBNL1 shows a temperature-dependent binding profile in cell extracts in the absence of ATP (data not shown), all recombinant protein binding studies were performed at 30°C to maximize binding while

minimizing RNA degradation. As anticipated, MBNL1 FL showed a high affinity in the filter binding assay for $(CUG)_{54}$ (K_d=5.3 ± 0.6 nM), but it also showed high affinities for the other RNAs examined including Tnnt3 5.45 (Tnnt5.45) (K_d=6.6 ± 0.5 nM) and $(CAG)_{54}$ (K_d=11.2 ± 1.5 nM) (Figure 4-3B). While there is no statistically significant difference between the K_ds of MBNL1 to $(CUG)_{54}$ and T5.45 (p=0.119), the K_d values for MBNL1 binding to $(CUG)_{54}$ and T5.45 are significantly different from the K_d of MBNL1 binding to $(CAG)_{54}$ (p=0.028 and p=0.002, respectively). Given the ~two-fold differences in the values of these K_ds, this result suggests that MBNL1 binds to $(CAG)_{54}$ with a different, yet similar, affinity compared to $(CUG)_{54}$ and T5.45. The similarities in the binding affinities of MBNL1 for $(CUG)_{54}$ and $(CAG)_{54}$ accounts for the prior observation that overexpression of either of these repeat RNAs in COS-M6 cells results in the formation of nuclear foci that colocalize with GFP-MBNL1 {Ho, 2005 #11}. In contrast, Tnnt3/T5.1 RNA, which did not crosslink to MBNL1 (Figure3-7B), bound poorly. Comparable affinities were obtained when MBNL1-RNA complexes were analyzed by gel shift analysis (Figure 4-3C). Incubation of full-length MBNL1 (MBNL1 FL) with Tnnt3/T5.45 generated several protein-RNA complexes resolved by the polyacrylamide gel whereas significant binding to Tnnt3/T5.1 was only detectable at 256 nM and only gel excluded complexes were observed. Similar complexes were also formed with MBNL1 N (data not shown). Incubation of MBNL1 with $(CUG)_{54}$ also resulted in the formation of several major complexes at protein concentrations (4-16 nM) near the K_d determined by filter binding. At higher MBNL1 FL concentrations (≥64 nM) the majority of the resulting MBNL1-$(CUG)_{54}$ complexes migrated at, or near, the top of the gel. The striking similarity in the binding affinities of MBNL1 for pathogenic and splicing precursor RNAs prompted us to re-examine the interaction of this splicing factor with CUG^{exp} RNA.

Visualization of MBNL1-CUG^{exp} Complexes

In a previous study, it was reported that the MBNL splicing antagonist, CUGBP1, binds to out-of-register ssCUG repeats at the base of CUG^{exp} RNA hairpins but not the A-form helical region while the dsRBD protein TRBP associates with the stem region {Michalowski, 1999 #111}. To confirm that MBNL1 is a stem-binding protein, electron microscopy (EM) was performed. Purified $(CUG)_{136}$ was examined following direct absorption to thin carbon foils, dehydration and tungsten shadowing. In contrast to ssRNA, $(CUG)_{136}$ RNA formed rod-like segments as described previously for $(CUG)_{130}$ (Figure 4-4A-C). To examine the structures of MBNL1-RNA complexes, MBNL1 was incubated with RNAs at two different RNA:protein molar ratios (1:2.5 or 1:10) and subsequently prepared for EM. In the presence of RNA, purified MBNL1 formed a ring-shaped structure with a prominent central cavity and for $(CUG)_{136}$-MBNL1 complexes incubated at a ratio of 1:2.5, ~70% of the RNAs were bound by one of these MBNL1 rings (Figure 4-4 D-F). At higher protein levels (1:10), free $(CUG)_{136}$ RNA was rarely detectable (6.4% of the RNAs in the field) while >90% of the RNAs were bound by two or more MBNL1 rings (Figure 4-4 G-H). Also shown is a representative field of dsCUG RNAs and MBNL1-$(CUG)_{136}$ complexes (Figure 4-4I, RNA:protein is 1:2.5). Although the majority of MBNL1 rings were associated with RNA under our RNA assembly conditions, a few free rings were visualized in the background in the absence of associated $(CUG)_{136}$ helices suggesting that MBNL1 may form a ring structure independent of RNA. We failed to visualize rings by negative staining, suggesting that this structure is disrupted by the acidic conditions of the negative staining protocol.

MBNL1 Self-interaction Mediated by the C-terminal Region

A number of studies have demonstrated that MBNL1 accumulates in ribonuclear foci together with pathogenic RNA {Mankodi, 2005 #124}{Mankodi, 2001 #38}{Jiang, 2004

#32}{Mankodi, 2003 #131}. While additional proteins might also bind to CUG^{exp} and $CCUG^{exp}$ RNAs, the number of ribonuclear foci in DM myoblasts declines significantly following loss of MBNL1, suggesting that this protein is required for the formation and/or maintenance of these unusual nuclear structures {Dansithong, 2005 #18}. Since many RNA-binding proteins function as components of large multi-subunit complexes and some of these proteins self-interact via their auxiliary or non-RNA binding regions{Moore, 2005 #564}{Cartegni, 1996 #565}, we tested the possibility that MBNL1 proteins self-associate using the yeast two hybrid system. Although the amino terminal region of MBNL1 contains all four CCCH motifs and is responsible for protein-RNA interactions {Kino, 2004 #21}, very little is known about the function of the C-terminal region. The MBNL1 FL protein showed strong homotypic interactions in this system (Figure 4-5A). Although MBNL FL failed to interact with the MBNL1 N-terminal region (1-264 amino acid residues), interactions between the full-length protein and C-terminal region (239-382 amino acid residues) were readily detectable. This C-terminal region does not contain any known RNA binding, or other, structural motifs.

To confirm that MBNL1 homotypic interactions occurred in a mammalian cell context, 293T cells were co-transfected with plasmids which expressed either V5-MBNL1 FL alone, V5-MBNL1 FL and MBNL1 FL-myc, or V5-MBNL1 FL and MBNL1 N-myc. Twenty-four hours following transfection, V5-tagged MBNL1 was immunopurified from cell lysates using an anti-V5 antibody and the precipitates were then immunoblotted using either anti-myc or anti-V5 antibodies. In agreement with the two hybrid analysis, the full-length V5 and myc-tagged proteins were associated *in vivo* while the N-terminal MBNL1 region (MBNL1 N-myc) failed to co-immunopurify with V5-MBNL1 FL (Figure 4-5B). Interactions between V5-MBNL1 FL and MBNL1 N-myc were resistant to RNase treatment since digestion of the cell lysate with RNase

A did not affect the amount of MBNL1 N-myc in the V5-MBNL1 FL immunoprecipitate. We conclude that MBNL1-MBNL1 interactions occur in vivo and these interactions are mediated by the C-terminal region.

MBNL1-RNA Complexes Display Distinct Stabilities

What leads to MBNL1 functional sequestration in DM if MBNL1 binds to pathogenic RNA and normal splicing targets with similar affinities? One possibility, as mentioned in the Introduction, is that the complexes formed between MBNL1 and pathogenic RNAs are more stable compared to those formed between MBNL1 and its normal RNA targets. A modified UV crosslinking assay was performed to test this hypothesis. A 2000 fold excess of non-labeled RNA was added to the nuclear extract either at the same time or 15 minutes following the addition of the uniformly radio-labeled RNAs to the reacions. RNA-protein mixtures were further incubated for 15 minutes, followed by UV irradiation and RNase A digestion. The percentage of protein remaining bound to the radiolabeled RNAs after the cold RNA challenge was calculated by comparing the amounts of crosslinked protein in lanes "concurrent" vs. lanes "after" where there was no cold competitor. CUGBP1 crosslinks to T5.45 strongly in the absence of cold T5.45 competitor. However, in the presence of an excessive amount of non-labeled T5.45 RNAs whether added concurrently or after the radiolabeled RNA, CUGBP1 crosslinking was considerably reduced to a similar extent, indicating a dynamic interaction between CUGBP1 and T5.45 RNA. However, MBNL1 displayed a much more static interaction pattern with all three RNAs tested, of which $(CUG)_{54}$ repeats was able to retain 100% of the MBNL1 crosslinking signal even after the addition of a 2000-fold excess of cold competitor, suggesting that the MBNL1-$(CUG)_{54}$ complex is not prone to dissociation once formed. The $(CAG)_{54}$ and T5.45 RNAs were less efficient in retaining MBNL1 compared to $(CUG)_{54}$, since ~40% and ~60% MBNL1 dissociated from $(CAG)_{54}$ and T5.45, respectively. In conclusion,

MBNL1 forms stable complexes with RNAs. However, the complexes formed between MBNL1 and CUG repeats display a considerably higher stability compared to those formed between MBNL1 and other RNAs.

Conclusions, Discussion and Future Work

MBNL1 Targets Similar Binding Motifs in Splicing Precursor and Pathogenic RNAs

Although there is considerable evidence for the MBNL loss-of-function model for DM pathogenesis, the molecular basis for MBNL1 sequestration by CUG^{exp} and $CCUG^{exp}$ RNAs has not been elucidated. Since mutant RNA expansions must compete with normal pre-mRNA, and possibly mRNA, binding sites for MBNL1 recruitment, effective MBNL1 sequestration might occur if the affinity of this protein for CUG^{exp} and $CCUG^{exp}$ RNAs is greater than for its normal splicing targets. The binding analysis presented here does not support this conjecture, since MBNL1 also possesses relatively high affinities for CAG^{exp} and Tnnt3 precursor RNAs. Indeed, these binding studies provide an explanation for the formation of MBNL1-containing ribonuclear foci in cells overexpressing CAG^{exp} {Ho, 2005 #11}. Additionally, mapping of a binding site to a stem-loop structure in Tnnt3 intron 8 just upstream of the fetal exon indicates that RNA recognition by MBNL proteins involves a common interaction mode for both pathogenic and normal pre-mRNAs: recognition of GC-rich hairpins containing pyrimidine mismatches.

What is the physiological significance of the binding preference of MBNL1 for RNA stem-loop structures? In this chapter, we provide evidence that MBNL1 is a splicing regulator which acts as an intronic splicing repressor by recognizing a stem-loop near the Tnnt3 fetal exon 3' splice site, possibly resulting in the stabilization of this secondary structure and interference with U2AF recruitment. Based on this observation, we speculate that MBNL1 recognizes similar intronic stem-loop structures adjacent to the 5' splice sites of other MBNL1-regulated fetal exons

to inhibit U1 snRNP recruitment or MBNL1 may stabilize interactions between introns flanking alternative exons resulting in RNA looping and increased fetal exon skipping.

The importance of RNA secondary structures in inherited disease and alternative splicing has been previously highlighted in frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP-17) which is caused by mutations in the *MAPT* gene encoding the microtubule-associated protein tau {Mankodi, 2003 #131}. Some FTDP-17 mutations destabilize a predicted stem-loop structure, which forms between the 3'end of exon 10 and the 5' end of the downstream intron, resulting in an increase in U1 snRNP recruitment and E10 inclusion. Interestingly, MAPT exon 10 skipping increases in the DM brain suggesting that MBNL1 promotes exon 10 inclusion during splicing {Kanadia, 2006 #157}. The MBNL1 binding preferences shown in this study suggest that this factor may also function as a splicing activator by recognizing RNA stem-loop structures in novel exonic and/or intronic splicing enhancers or by blocking splicing silencer elements by stabilizing overlapping RNA secondary structures.

Another interesting question arose when MBNL1 binding site on Tnnt3 was mapped to an 18-nt stemloop structure. Previous study showed that MBNL1 in HeLa nuclear extract crosslinks strongly to $(CUG)_{74}$ and $(CUG)_{97}$ but not to $(CUG)_{<20}$ {Kino, 2004 #21}. Based on this observation, we proposed that below a certain length threshold (<20 repeats) the dsCUG helix was unstable in the cell extract and ssCUG was not a binding site for MBNL1. The results reported here support that proposal and demonstrate that MBNL1 is primarily a dsRNA binding protein which recognizes relatively short GC-rich hairpins if the overall RNA secondary structure is stabilized by additional sequence interactions. This conclusion provides a plausible resolution for the apparently conflicting results that overexpression of a GFP-DMPK 3'-UTR

$(CTG)_5$ transgene results in a DM phenotype {Mahadevan, 2006 #156} while mice expressing the *HSA*^{SR} (human skeletal α-actin containing a 3'-UTR with five CTG repeats) transgene are normal {Mankodi, 2000 #16}. For the *DMPK* 3'-UTR, the $(CTG)_5$ repeat is predicted to interact with an upstream region to form a GC-rich stem interrupted by several U-U and C-C mismatches while this repeat in the HSA 3'-UTR is located in sequential unpaired loops. It is possible that this structural arrangement in the DMPK 3'-UTR promotes MBNL1 sequestration when GFP-DMPK 3'-UTR $(CTG)_5$ RNA is overexpressed in transgenic mice.

MBNL1 Rings: Interactions with CUG^{exp} RNA and Potential Mechanism for MBNL1 Functional Sequestration

Filter binding and gel shift assays indicated that the affinity of MBNL1 for Tnnt3/T5.45 is similar to those of $(CUG)_{54}$ and $(CAG)_{54}$ in cell-free binding reactions. While overexpression of either CUG or CAG repeats induces the formation of nuclear foci in cell culture, it is interesting to note that there was no detectable RNA foci formation following Tnnt3 minigene overexpression (data not shown). This observation argues that high affinity MBNL1-RNA interactions together with abundant expression of MBNL1 target RNAs is not sufficient for ribonuclear foci formation. Of course, MBNL1 may be cleared from splicing target, but not pathogenic, RNAs during RNA processing and nuclear export or unusual interactions between MBNL proteins and CUG^{exp} RNA might drive foci formation. In support of the latter possibility, we provide EM evidence that MBNL1 forms a tandem ring structure when bound to CUG^{exp} RNA. The size of the rings is uniform with a diameter of ~18 nm and since the MBNL1 isoform employed for these studies is 41 kDa, the ring structure must be an oligomeric complex. At a protein:RNA ratio of 2.5:1, most CUG^{exp} helices were bound by a single ring while the majority of these hairpins were bound by at least two rings at a higher protein:RNA ratio. It is not clear if the MBNL1 ring contains a hole but if a central cavity exists it might allow threading of dsRNA.

When multiple rings were bound to a single RNA molecule they appeared to be tandemly stacked suggesting either a preferential ring-loading site or potential ring-ring interactions. The latter possibility is supported by the finding that MBNL1 self-interacts via its C-terminal region both in the yeast two hybrid system and in mammalian cells. In contrast, the MBNL1 N-terminal region encompassing the CCCH RNA binding motifs fails to interact with full length MBNL1 although RNA-binding activity is comparable to the full-length MBNL1 (data not shown). When MBNL1 N proteins were incubated with CUG_{136}, similar ring structures were also detectable by EM although ring size was less uniform and there was considerably less stacking of MBNL1 N rings compared to MBNL1 FL (data not shown). A similar situation has been noted for the *E.coli* protein Hfq (Host factor 1) which is a single-strand RNA binding protein involved in the translational regulation and stability of several RNAs (37). As visualized in the EM by negative staining, Hfq forms hexameric rings. Similar to MBNL1, the RNA binding activity of Hfq resides in the N-terminal region and rings are still formed by a C-terminal truncated protein although they are less stable.

Does MBNL1 exist as a ring structure when bound to its normal RNA splicing targets? To address this question, EM analysis was performed on the MBNL1-Tnnt3/T5.45 complex. Although MBNL1 rings were observed, the result was inconclusive due to the difficulty in visualizing small and partially single-stranded Tnnt3/T5.45 RNA by EM. However, MBNL1 forms large complexes with both CUG and Tnnt3/T5.45 RNAs (Figure 4-4 C) so it is possible that MBNL1 forms a ring structure when bound to splicing regulatory sites.

What leads to MBNL1 functional sequestration on pathogenic RNAs? Based on the observations in this study, a model is proposed as illustrated in Figure 4-7. MBNL1 forms the ring structure around G-C rich stemloop RNA. The carboxyl terminus of MBNL1 is not

essential for ring formation, yet it is involved in stabilizing the stacking of the MBNL1 rings. When MBNL1 binds to expanded CUG or CCUG RNAs, multiple ring structures could be loaded onto the long RNA stemloop. With the C-terminus stabilizing the MBNL1-MBNL1 interactions, the structure is thermodynamically stable and resistant to disruption. On the other hand, when MBNL1 binds to its splicing targets, only one ring structure could be accommodated due to the short length of the binding sites. In the absence of the reinforcing stacking interactions, the rings are prone to dissociate from the RNA. Thus, the higher stabilities of MBNL1-pathogenic RNA complexes lead to the accumulation of MBNL1 on pathogenic RNA. Due to the extremely low off –rate of MBNL1 molecules bound to pathogenic RNAs (Figure 4-6), the CUG or CCUG-bound MBNL1 molecules are inaccessible to the normal splicing targets. Interestingly, based on the observation that $(CUG)_{136}$ RNA could accommodate 3-4 MBNL1 rings, the footprint of each MBNL1 ring is estimated to be 34-45 CUG repeats, which is slightly lower than the DM1 pathogenic repeat range (>50), supporting the hypothesis that MBNL1 ring stacking is required for its functional sequestration.

As demonstrated in Figure 4-6, 54 CAG repeats retain MBNL1 much less than CUG repeats, which could explain the results by Ho and colleagues that $(CAG)_{960}$ failed to promote embryonic splicing patterns for *cTNT* and *IR* as $(CUG)_{960}$ did {Ho, 2005 #11}. I think that structural differences between CAG and CUG repeats might underlie their varied ability to functionally titrate MBNL1. The U-U mismatch in CUG repeats do not to cause any distortion in the RNA sugar-phosphate backbone {Mooers, 2005 #227}. Although no structural study has been carried out for CAG repeat RNA, it is possible that the bulged purine mismatches in CAG lead to a different RNA helix structure and block MBNL1 ring stacking.

Future Work

The hypothesis that the extreme stability of the MBNL1-pathogenic RNA complex leads to MBNL1 functional sequestration requires further experimental support. Studies such as a plasmon resonance analysis need to be carried out to confirm the distinct stabilities of MBNL1-RNA complexes seen in the competition assay. If indeed the difference in stabilities is confirmed, it would be interesting to ask what factors contribute to the enhanced stability of the MBNL1-pathogenic RNA complex. Is this stability mediated by MBNL1 ring stacking as illustrated in Figure 4-7? If so, once the length of the CUG repeats is reduced below a certain threshold where ring stacking is not allowed, would MBNL1-CUG complexes display a similar stability compared to the MBNL1-Tnnt3/T5.45 complex? Based on the considerably weaker stacking ability of MBNL1 N-terminual rings compared to MBNL1 full length observed in the EM, this model would predict that the MBNL1 N-terminus is unable to form the ultra-stable complex with long CUG repeats as the full-length protein does. These proposed experiments are essential for validating our proposed hypothesis.

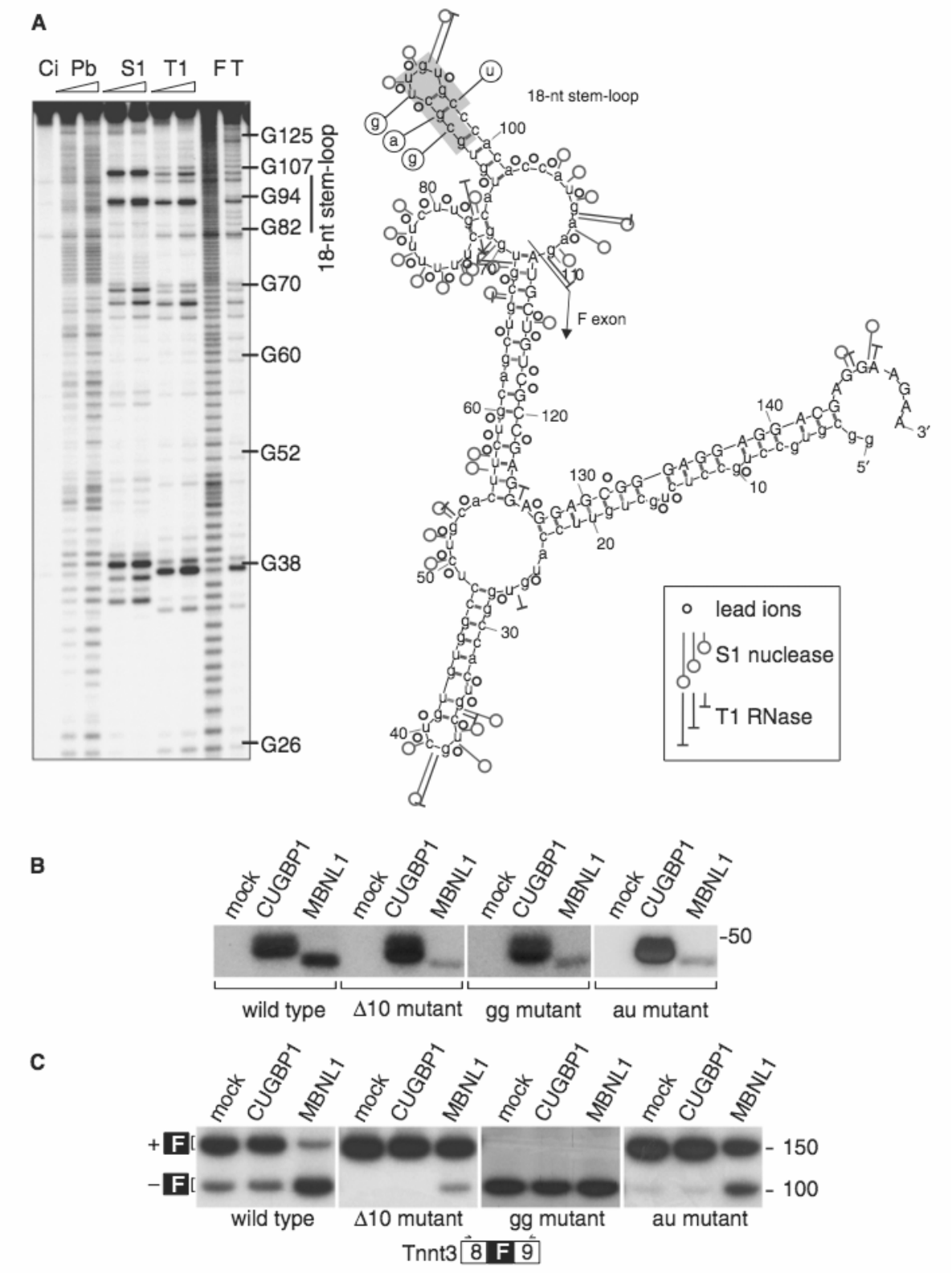
A
Ci Pb S1 T1 F T
G125
G107
G94
G82
G70
G60
G52
G38
G26
18-nt stem-loop
18-nt stem-loop
F exon
lead ions
S1 nuclease
T1 RNase
B
mock CUGBP1 MBNL1 mock CUGBP1 MBNL1 mock CUGBP1 MBNL1 mock CUGBP1 MBNL1
-50
wild type
Δ10 mutant
gg mutant
au mutant
C
mock CUGBP1 MBNL1 mock CUGBP1 MBNL1 mock CUGBP1 MBNL1 mock CUGBP1 MBNL1
+F
−F
- 150
- 100
wild type
Δ10 mutant
gg mutant
au mutant
Tnnt3 8 F 9

Figure 4-1 MBNL1 recognizes a RNA hairpin upstream of the Tnnt3 fetal exon. **A**) Cleavage pattern (left) of the 5'-end labeled Tnnt3 151-nt transcript encompassing the fetal exon 3' splice site (110-nt of intron 8, 41-nt of fetal exon) obtained with use of three structure probes. Lanes are: Ci, incubation control or no probe added; Pb, lead ions (0.25, 0.5 mM); S1, S1 nuclease (1, 2 U/µl and 1 mM $ZnCl_2$ was present in each reaction); T1, RNase T1 (0.5, 1 U/µl); F, formamide (statistical ladder); T, guanine-specific ladder. The sequences forming the 18-nt stem-loop structure are also indicated. Also illustrated (right) is the proposed secondary structure model of the 151-nt transcript. The cleavage sites are indicated for each probe used and the figure inset shows the probe designations and cleavage intensity classification. The fetal exon sequence is marked in upper case and intron 8 in lower case. The positions of two G substitutions in the 18-nt stem-loop are also indicated. **B**) Photocrosslinking analysis showing reduced MBNL1, but not CUGBP1, binding to the Tnnt3 Δ10 and gg mutants in contrast to wild type RNA. Photocrosslinking analysis was performed as described in Fig. 1B except only MBNL1 FL (MBNL1) protein was used. **C**) Tnnt3 fetal exon skipping is impaired in the Δ10 Tnnt3 T5.45 mutant compared to wild type while fetal exon inclusion is eliminated in the gg double mutant. C2C12 cells were co-transfected with either a wild type, Tnnt3 Δ10 or gg double point mutant splicing reporter plasmid and a protein expression plasmid for either CUGBP1mycHis or MBNL1mycHis (full-length protein only). ^{32}P-labeled splicing products, which included (+F) or excluded (-F) the Tnnt3 fetal exon (black box), were detected by RT-PCR, using primers positioned in Tnnt3 exons 8 and 9 (open boxes with arrows), followed by gel electrophoresis.

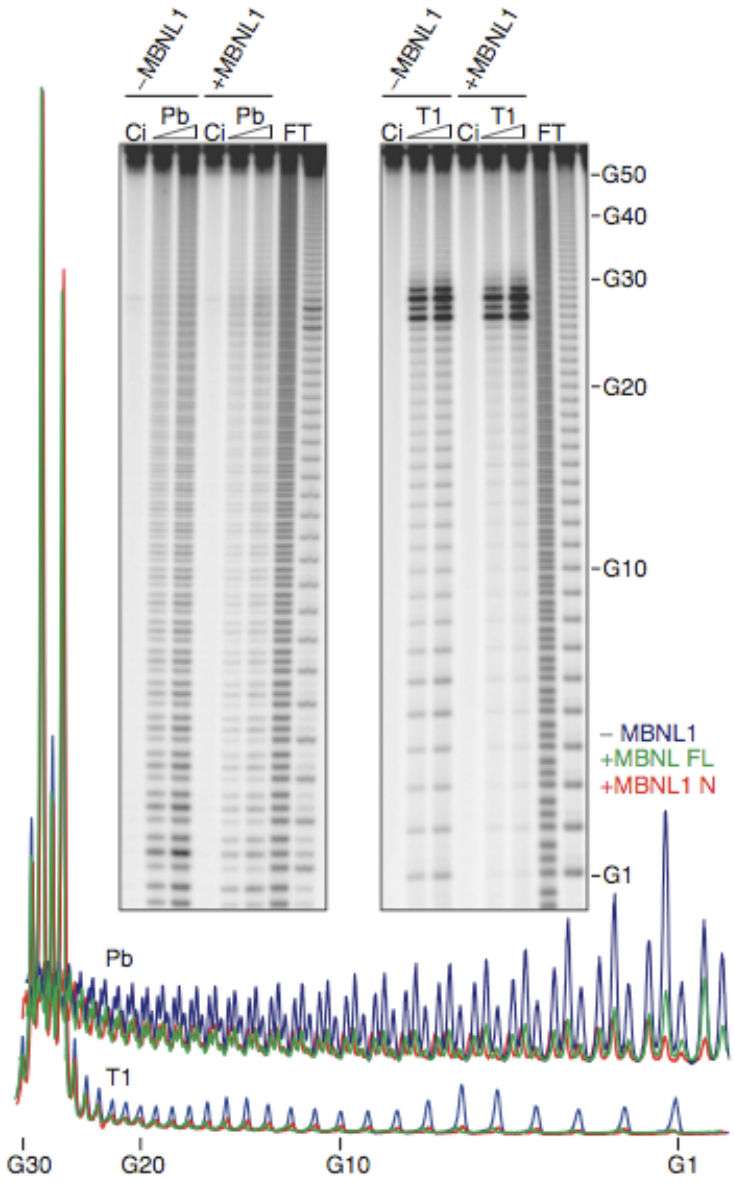

Figure 4-2 MBNL1 binds throughout the dsCUG stem. Structural analysis of 5'-end labeled $(CUG)_{54}$ transcript (20 nM) either in the presence (+) or absence (-) of 500 nM recombinant MBNL1 and either lead ion (0.25 or 0.5 mM, lanes Pb) or ribonuclease T1 (0.5 or 1 units/µl, lanes T1); incubation control (no probe added, lane Ci); formamide (lane F); guanine-specific ladder obtained with RNase T1 compete digestion (lane T). The positions of selected G residues are shown along the T1 ladder (G-residues of the corresponding CUG repeat are indicated). Below the gel panels is the color-coded densitometric analysis (ImageQuant) of cleavage patterns for the CUG stem obtained with lead ion (Pb) and T1 ribonuclease (T1) in the presence of MBNL1 FL (green), MBNL1 N (red) or in the absence of protein (blue).

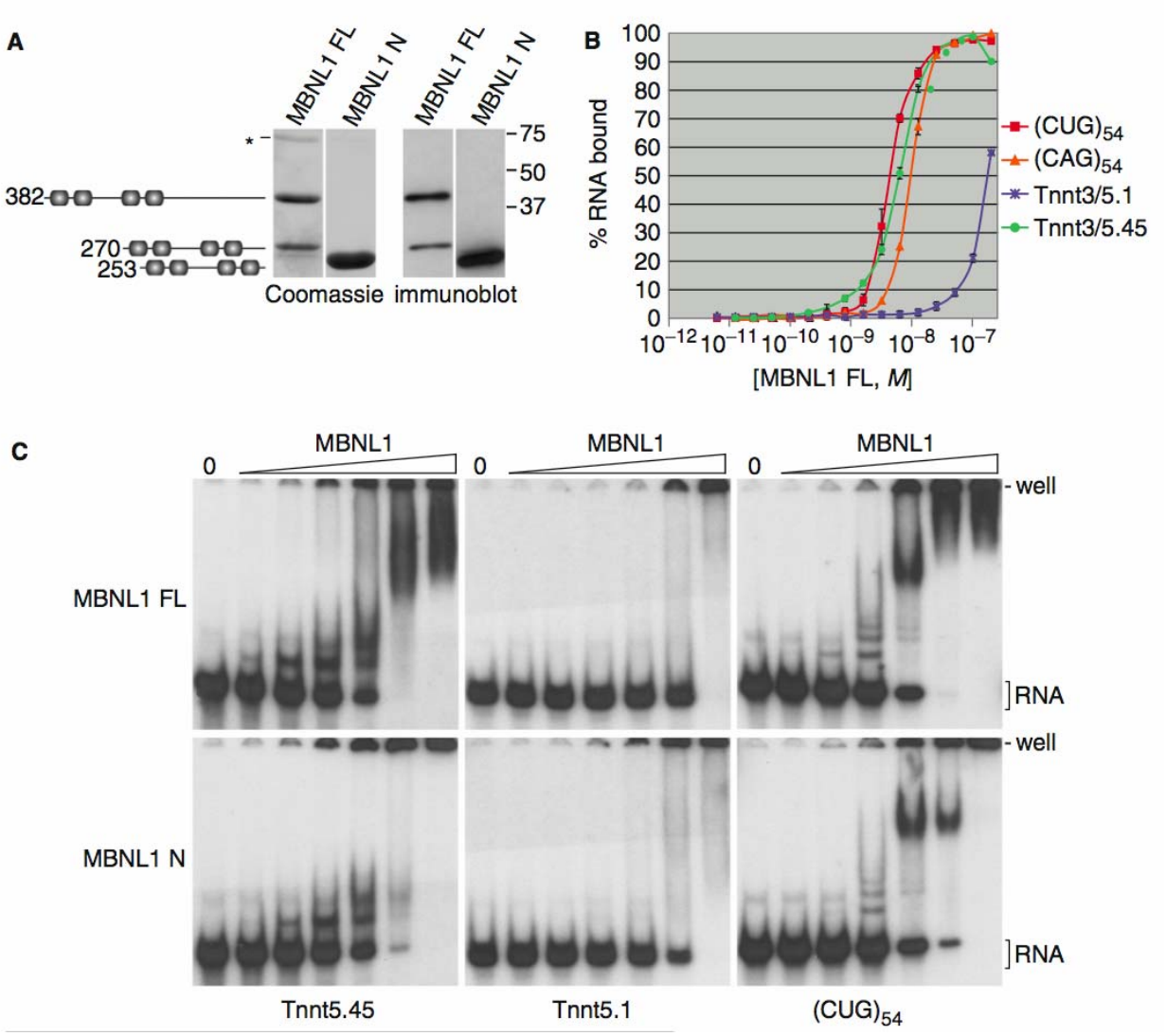

Figure 4-3 MBNL1 binds to pathogenic and splicing precursor RNAs with similar affinities. **A**) Purified recombinant MBNL1 proteins. Coomassie-stained gels (left panels) and immunoblots (right panels) of either full length or N-terminal proteins. Illustration shows the primary structures (line; C_3H motifs are shown as shaded boxes) indicating the full-length (382 aa) and N-terminal (253 aa) MBNL1 proteins together with a C-terminal truncated protein (270 aa) generated during expression in *E. coli* and an unknown 75 kDa protein (asterisk). **B**) Nitrocellulose filter binding analysis of MBNL1 FL binding to $(CUG)_{54}$ (red square), $(CAG)_{54}$ (orange triangle), Tnnt3 5.1(T5.1, blue cross) and Tnnt3 5.45 (T5.45, green circle) RNAs (see Figure 3-7). **C**) Electrophoretic mobility shift analysis of MBNL1 FL binding to Tnnt3 5.45, 5.1 or $(CUG)_{54}$. The positions of the free RNA (bracket), the gel origin (well) and MBNL1 FL concentrations (triangle, lanes are 0, 0.25, 1, 4, 16, 64, 256 nM) are indicated.

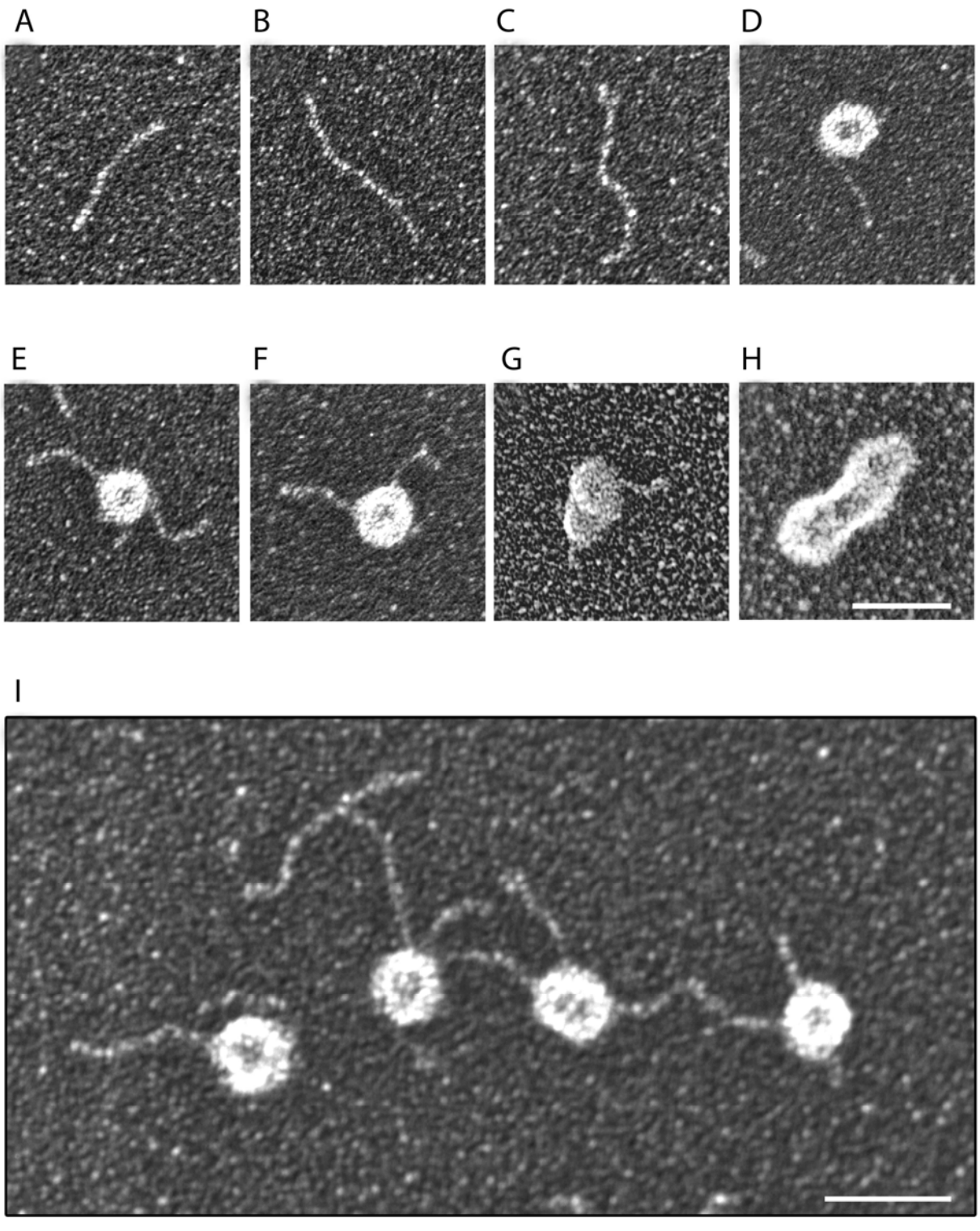
A
B
C
D
E
F
G
H
I

Figure 4-4 Visualization of dsCUG and MBNL1-dsCUG complexes. Electron microscopy of either free $(CUG)_{136}$ RNAs (**A-C**) or MBNL1-$(CUG)_{136}$ complexes (**D-I**) at a protein:RNA ratio of either 2.5:1 (**D-F, I**) or 10:1 (**G-H**). As reported previously, purified dsCUG RNAs, which were directly adsorbed onto thin carbon foils followed by dehydration and rotary shadow casting with tungsten, are elongated rod-shaped structures (23). For analysis of complexes, MBNL1 protein and $(CUG)_{136}$ RNA were incubated together at 30°C, fixed with glutaraldehyde, passed over a gel filtration column and adsorbed on copper grids for rotary shadowing. Under these conditions, MBNL1 has a distinct ring-shaped structure. Size bars (white line) are 40 nm (**H**) or 50 nM (**I**).

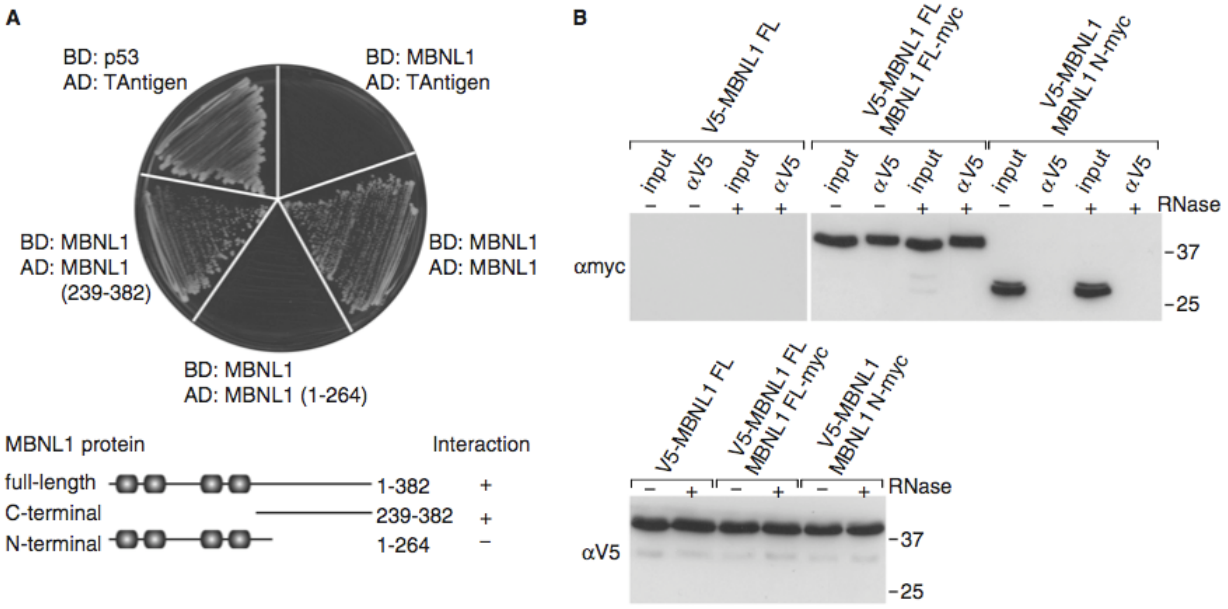

Figure 4-5 Self-association of MBNL1 proteins is mediated by the C-terminal region. **A**) Two hybrid analysis using yeast strains transformed with the following binding domain (BD) and activation domain (AD) plasmids: 1) GAL4 DNA binding domain (BD) plasmids with either p53 (activation control) or full-length MBNL1(MBNL1); 2) activation domain plasmids with either T-antigen (activation control), MBNL1 (residues 1-382), MBNL1 1-264 (N-terminal region) or MBNL 239-382 (C-terminal region). Functional interactions between the proteins expressed from the BD and AD plasmids results in growth on the Trp^- Leu^- His^- selection plate. **B**) Co-immunopurfication of MBNL1 requires the C-terminal region. HEK293T cells were co-transfected for 24 h with plasmids expressing tagged versions of either full-length or N-terminal MBNL1 (V5-MBNL1 FL alone, co-transfected V5-MBNL1 FL and MBNL1 FL-myc, co-transfected V5-MBNL1 FL and MBNL1 N-myc). Cell lysates were prepared and the V5-MBNL1 FL protein immunopurified (αV5) followed by SDS-PAGE (50% of IP sample) and immunodetection of MBNL1 FL-myc or MBNL1 N-myc using mAb 9E10 (top panel, input lanes represent 2.5% of total IP) or V5-MBNL1 FL (bottom panel, only the input lanes are shown).

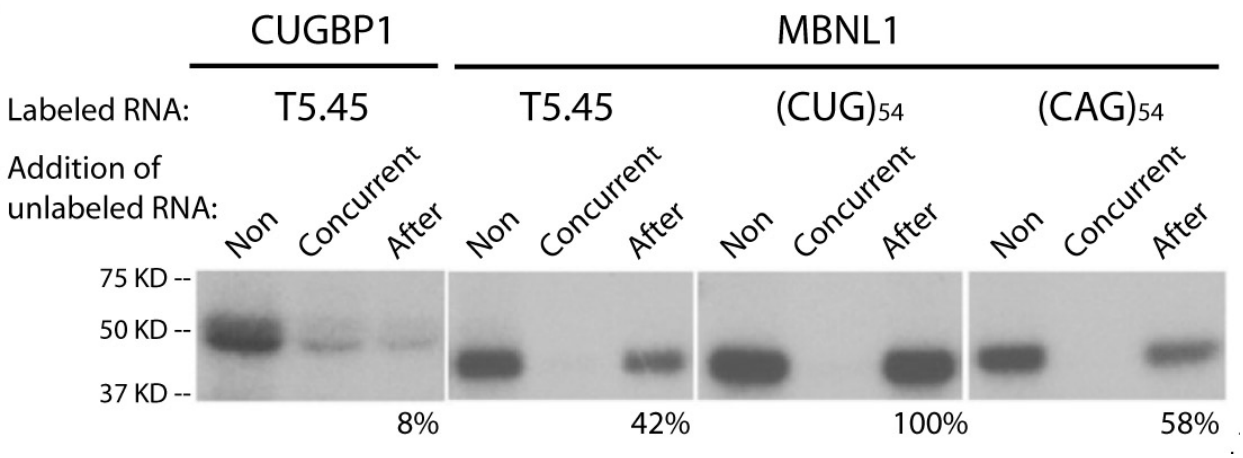

Figure 4-6 MBNL1-RNA complexes display distinct stabilities. UV crosslinking assays were performed in the absence of competitors (non) or in the presence of a 2000-fold excess of non-labdeled RNAs. The cold RNAs were either added concurrently with the labeled RNAs (Concurrent) or added after the labeled RNAs has been incubated with the nuclear extract for 15 minutes (After). Reactions were incubated for another 15 min followed by UV crosslinking and RNase A digestion.

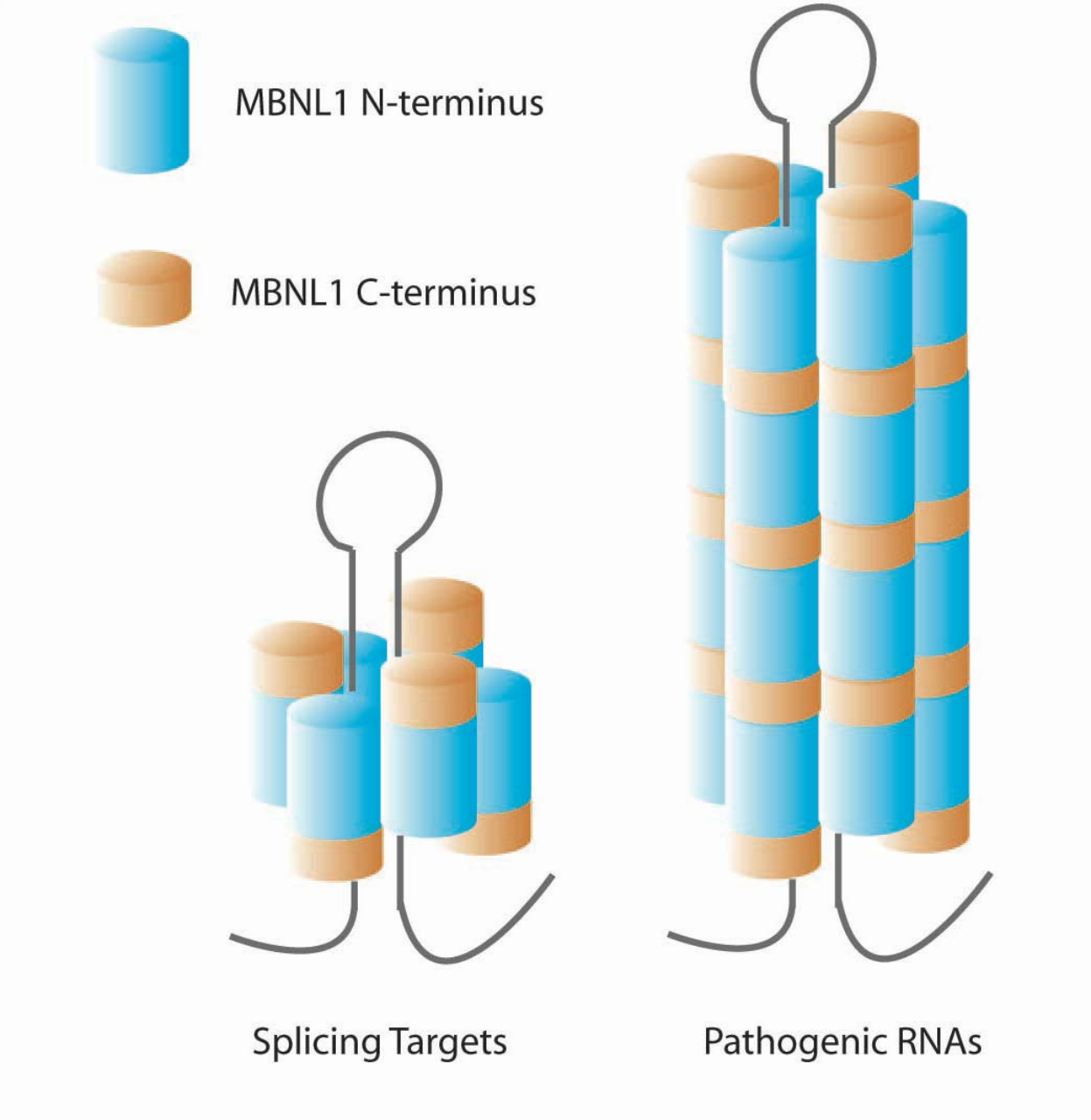

Figure 4-7 Model of complexes formation between MBNL1 and splicing targets versus pathogenic RNAs. MBNL1 forms oligomeric ring structure upon binding to RNA. Normal splicing targets harbor short hairpin structure which is long enough to accommodate only one MBNL1 ring. One the other hand, pathogenic RNAs contain tandem MBNL1 binding sites. Stacking of MBNL1 rings mediated by its C-terminus stabilizes the complexes formed between MBNL1 and pathogenic RNAs.

CHAPTER 5
DISCUSSION: INVOLVEMENT OF MUSCLEBLIND-LIKE PROTEINS IN RNA-MEDIATED DISEASES AND POTENTIAL THERAPEUTIC APPROACHES

DM1 and DM2

By using CUG repeats as an example for pathogenic RNAs, my study mainly focuses on MBNL1 involvement in DM1 pathogenesis. The other type of myotonic dystrophy, DM2, is caused by a CCTG expansion in the first intron of *ZNF9*. Compared to DM1, DM2 is a milder disease in that the age of onset is generally much older, and the symptoms are less severe. However, the sizes of CCTG expansions in DM2 are usually much larger than those of the CTG expansions in DM1. In addition, the expression of *ZNF9* is ubiquitous and about ten fold higher than that of *DMPK* (153). If MBNL1 sequestration by CUG or CCUG repeats is the major mechanism underlying both DM1 and DM2 pathogenesis, what might account for the milder symptoms in DM2 given the larger and more abundant CCUG RNAs? Several possibilities exist. First, MBNL1 may have a higher affinity for CUG repeats compared to CCUG repeats. Yet, to the contrary, MBNL1 displayed stronger interaction for CCUG repeats than CUG repeats in yeast three-hybrid system (116). This observation was further supported by our preliminary filter binding result using an interrupted (CCUG)46 tract (data not shown). Surprisingly, the K_d of MBNL1 for this interrupted CCUG repeat was an order of magnitude lower than that of a CUG repeat of comparable length. Thus MBNL1 appears to have a higher affinity for CCUG repeats.

Another possibility is that CCUG RNA repeats possess a reduced ability to retain MBNL1 protein compared to CUG repeats. The stemloop predicted by Mfold which is formed by CCUG repeat RNA has two, instead of one, pyrimidine mismatches between each G-C Watson-Crick basepair. This additional pyrimidine mismatch could potentially affect the RNA sugar-phosphate backbone conformation and/or surface electrostatic potential, which could result in a

distinct MBNL1 binding mode (*65*). Our preliminary EM result suggests that MBNL1 rings formed on CCUG repeat are smaller in size compared to the ones formed on CUG repeats. Moreover, there was considerably less ring stacking on CCUG (data not shown). It would be interesting to test in the commitment assay whether expanded CCUG retains MBNL1 to a lesser extent than CUG repeats, and if CCUG display a reduced ability to affect MBNL1-mediated splicing switches compared to CUG repeats. If these results turn out as predicted, it could provide a mechanistic basis for the difference in disease severity between DM1 and DM2.

Congenital and Adult-Onset DM1

Congenital DM patients present a very distinct array of symptoms compared to adult-onset DM1 patients, including hypotonia and severe mental retardation. Notably, myotonia, which is 100% penetrant among adult-onset DM patients, is absent in CDM patients. The fact that CDM patients recover some of the aforementioned symptoms during their childhood, and eventually develop adult-onset symptoms, suggests a different pathogenesis mechanism for CDM. We have shown that five known MBNL1 targets all undergo splicing switches during a similar postnatal time window in mice (122). If only those functions of MBNL1 involved in promoting adult splicing patterns were impaired in CDM patients, severe consequences should not be expected in newborns, since MBNL1 regulated splicing switches have not yet been activated. Why, then, are congenital DM patients with large CTG expansions affected in infancy? A number of genes, including fibronectin, agrin, and alpha1(X1) collagen chain, have been shown to undergo prenatal splicing switches during embryogenesis (161-163). One possibility is that MBNL3 is required for certain prenatal splicing switches. Indeed, *Mbnl3* expression levels peak in the early stages of mouse embryonic development followed by a gradual decline during the later stages (*80*). Whole-mount *in situ* hybridization revealed high levels of Mbnl3 expression in the developing head region and limb buds in 9.5 dpc mouse embryos (*80*). If MBNL3 is important

for regulating prenatal alternative splicing events, co-expression of mutant DMPK transcripts in these regions may result in MBNL3 titration, and subsequent failure to complete certain prenatal splicing switches. Such failures may lead to the brain and muscle abnormalities observed for CDM patients. Functional sequestration of MBNL3 may require the accumulation of large amounts of mutant DMPK transcripts in the nucleus. Increased *DMPK* expression has been shown in CDM skeletal muscle, possibly through a CTCF-mediated chromatin insulator (164,165). Thus, compared to adult-onset DM patients, higher levels of mutant RNAs combined with longer CTG tract in CDM can sequester more MBNL3 protein, potentially sufficient to cause MBNL3 loss-of-function. To test this hypothesis, *Mbnl3* knockout mice are being generated currently. If mice lacking Mbnl3 recapitulate the CDM brain and muscle defects, it could provide a mechanistic explanation for the CDM pathogenesis. Meanwhile, Mbnl3 binding targets can be identified using the CLIP method. Subsequent studies could be done to test whether these CLIP-tagged pre-mRNAs undergo prenatal splicing switches regulated by MBNL3.

Potential MBNL Involvement in Other RNA-Mediated Diseases

Although first identified in the etiology of myotonic dystrophy, RNA-mediated disease pathogenesis is probably not unique to DM. In the past few years, researchers have identified a growing number of other dominantly inherited diseases possibly caused by toxic gain-of-function at the RNA level. Similar to myotonic dystrophy, toxic RNAs in these diseases often result from transcription of microsatellite expansions in non-coding regions. Interestingly, many of these disease-causing microsatellite expansions are also GC rich. Therefore, the transcribed RNAs in many cases are predicted to form GC-rich double-stranded stemloop structures with interspaced mismatches, reminiscent of the structures formed by $(CUG)_n$ and $(CCUG)_n$ RNAs. Based on our result that MBNL1 prefers GC rich stemloops with pyrimidine mismatches, these

mutant RNAs could potentially sequester MBNL proteins in a manner similar to that described in DM.

Spinocerebellar Ataxia Type 8 (SCA8)

SCA8 is a slow progressive form of cerebellar ataxia, characterized by gait and limb ataxia, nystagmus and dysarthria (166,167). The SCA8 mutation was identified to be a CTG expansion residing in a non-coding gene on human chromosome 13q21 (166). It has been reported that in the general population, more than 99% of the alleles have 16-37 CTA/CTG combined repeats. Alleles with 107-127 CTG repeats are associated with SCA8 patients (166). However the inheritance pattern of SCA8 is complex displaying incomplete penetrance, which may due to the size of CTA tract preceding the CTG expansion or different *SCA8* expression levels among patients (166-172).

Although the mechanism of SCA8 pathogenesis remains an area of debate, one hypothesis suggests that the CUG containing RNA is toxic resulting in impaired cellular function. This idea has gained experimental support from transgenic mouse and fly models. Transgenic mice expressing mutant *SCA8* BAC developed a progressive neurological phenotype, which was not present in mice expressing the normal *SCA8* transgene, suggesting that the expanded SCA8 RNA is pathogenic (173). In a *Drosophia* model, retinal expression of both normal (9 CTG) and expanded (107 CTG) human SCA8 non-coding RNAs resulted in a rough-eye phenotype (131). Interestingly, a genetic modifier screen uncovered mutations in *Drosophila* muscleblind gene which enhanced the rough-eye phenotype caused by *SCA8* expression in a CTG length dependent manner suggesting possible interactions between muscleblind proteins and SCA8 mutant transcripts (131). Further studies such as rescuing SCA8 transgenic mice by MBNL overexpression could be performed to determine whether MBNL1 is involved in SCA8 pathogenesis.

Huntington's Disease-like 2 (HDL2)

HDL2 is an autosomal dominant, progressive neurodegenerative disease characterized by chorea, dystonia, rigidity, bradykinesia, psychiatric syndromes, dementia, and an inevitable decline to death (174-180). HDL2 is caused by a CTG expansion in the gene junctophilin-3 (*JPH3*). Depending on the alternative splicing of the pre-mRNA, CUG repeats may encode polyleucine or polyalanine, or alternatively reside in the 3'-UTR (178). CTG repeat lengths in HDL2 display fairly tight normal versus pathogenic ranges with 6-28 triplets in normal individuals whereas 40-59 triplets are found in HDL2 patients (179). RNA foci have been detected in HDL2 frontal cortex using riboprobes specific for CUG repeats as well as junctophilin-3 mRNA (133). These RNA foci resemble those found in DM in terms of size, nuclear location, and, importantly, MBNL1 colocalization (133). Assayed by immunofluorescence, nuclear MBNL1 levels were reduced by an average of ~50% in foci containing cells, possibly leading to the aberrant splicing of MBNL1 targets in the affected brain region (133). These results suggest molecular parallels between DM and HDL2 pathogenesis.

Fragile X Tremor Ataxia Syndrome (FXTAS)

FXTAS is a late-onset neurodegenerative disorder with core features of action tremor and cerebellar gait ataxia (181). FXTAS is caused by a CGG expansion in the *Fragile X Mental Retardation 1 (FMR1)* gene (182). While full mutations of over 200 CGG repeats cause Fragile X Syndrome by silencing *FMR1* transcription, CGG expansions in the permutation range of 55-200 repeats lead to FXTAS, potentially through a proposed RNA gain-of-function mechanism (182,183). Results from transgenic mouse and fly models expressing CGG repeats of premutation length lend support to the RNA-mediated pathogenesis model. Both mice and flies develop neurodegenerative phenotypes upon rCGG repeat expression (184,185).

Consistent with the RNA gain-of-function model, rCGG repeats were found in ubiquitin-positive intranuclear inclusions in FXTAS brains (183). These inclusions have been recently purified from post-mortem brain tissue of FXTAS patients by fluorescent-activated cell sorting (FACS) (134). Mass spectrometric analysis of the protein composition of these inclusions revealed more than twenty inclusion associated proteins, among which are structural proteins such as lamin A/C and neurofilament 3, stress-related proteins, and RNA-binding proteins including hnRNP A2/B1 and MBNL1. Immunofluorescent staining further confirmed the localization of MBNL1 to the ubiquitin positive nuclear inclusions (134). Although MBNL1 does not bind to rCGG repeats with high affinity in filter binding assay (data not shown), considering the complexity of protein components in FXTAS inclusions, it remains possible that MBNL1 is indirectly recruited to the rCGG inclusions via other protein factors. Whether MBNL1 recruitment to rCGG foci plays a causative or secondary role in FXTAS pathogenesis requires careful examination using both genetic and biochemical approaches.

Implications for Potential DM Therapeutic Strategies

The ultimate goal of studying DM pathogenesis is to not only understand what goes wrong in the disease, but also find ways to reverse disease progression. MBNL loss-of-function plays a causative role in DM pathogenesis. Thus, a straightforward therapeutic strategy would be to supplement DM tissues with exogenous MBNL1 protein to help maintain adult splicing patterns. In fact, recombinant adeno-associated virus (rAAV) mediated rescuing experiment has been performed in the HSA^{LR} poly(CUG) mouse model for DM (159). Viral-mediated MBNL1 expression was induced in tibialis anterior (TA) muscle by injecting rAAV. Myotonia in the injected TA muscle was reversed up to 43 weeks post-injection, and aberrant splicing of several MBNL1 targets was successfully rescued. While effective for local applications, systemic rescue

of DM phenotypes using rAAV-mediated gene therapy requires a large amount of viral vector, which may pose an obstacle for potential clinical applications of this approach.

Another therapeutic strategy would be to disrupt the interaction between MBNL and pathogenic RNAs, thus enabling MBNL proteins to bind to their physiological targets and promote their adult splicing patterns. There are at least two ways to achieve this goal. The first strategy is based on the observation that MBNL1 has low affinities for perfectly base-paired dsRNAs (186). Thus, short stabilized $(CAG)_n$ or $(CAGG)_n$ could be introduced into DM cells and transform the CUG or CCUG hairpin into perfectly paired CAG•CUG or CCUG •CAGG which would no longer be able to titrate MBNL. Several commercially available nucleic acid modifications, such as the locked-nucleic acid (LNA), could be used to enhance the thermostability of the CAG•CUG or CCUG •CAGG hybrid. The toxicity of these modified nucleic acids requires careful assessment and their delivery methods await further improvements. With the rapid development of small molecule drug screens, thousands of chemicals could be screened for their abilities to inhibit MBNL binding to CUG or CCUG repeats, which could provide another promising avenue to achieving a molecular therapy for DM. However, a key to the successful clinical applications of these drugs will rely on their abilities to preserve the interactions between MBNL1 and normal RNA targets while preferentially disrupting the interactions between MBNL1 and pathogenic RNAs. Results from this study suggest that MBNL1 interacts with pathogenic and normal RNA targets in a very similar manner, in terms of recognition properties and binding affinities. Thus, it is highly likely that small molecules which disrupt MBNL1-CUG or CCUG binding would also inhibit MBNL1 binding to its physiological targets. A different screening strategy is proposed here based on the model illustrated in Figure 4-7. If MBNL1 C-terminus-mediated ring stacking is essential for MBNL1 functional

sequestration, chemicals could be screened for their abilities to disrupt MBNL1 C-terminal self-interactions. Since the MBNL1 N-terminus alone promotes adult splicing patterns to the same extent as full-length MBNL1 (Figure 3-6), such chemical agents would probably not affect MBNL1 regulated splicing events. However, based on the model in Figure 4-7, loss of MBNL1 C-terminal interactions would decrease ring stacking, which would lead to reduced stability of MBNL1-pathogenic RNA complexes, and subsequently alleviate MBNL1 functional sequestration. One important consideration of this strategy is the toxicity of naked CUG and CCUG repeats. Once stripped of MBNL proteins, CUG and CCUG repeat RNAs may be able to interfere with other cellular processes through interactions with protein or RNA factors with relatively high affinities. This possibility can be tested by crossing poly(CUG) or poly(CCUG) transgenics with $Mbnl1^{-/-}$ $Mbnl2^{-/-}$ double knockouts and assaying for additional phenotypes compared to the transgenic mice.

APPENDIX
LIST OF PRIMERS

Name	Sequence
MSS1163	GCGCCCGGGGATGGCTGTTAGTGTCACACCAATT
MSS1164	GCGCCCGGGTCCCCCTGCACATTTACAAGCC
MSS1165	GCGGGATCCTGGCTGCAGCCTGTGCAGCTG
MSS1166	GCGGGATCCACATCTGGGTAACATACTTGTGGC
MSS1865	TGAGACTAGGTGGTAGAAGGCAATGGAAGG
MSS1866	CTTCCTCGTCCTCCTCCCGCTC
MSS1879	AAAGCTTGGCTGTGAGGAGGGAGGA
MSS1884	CCTATGGGAAGGTGTAGGAGCTGCCC
MSS1938	GCTGCAATAAACAAGTTCTGCTTT
MSS1949	GGCGAATTCAGGAAGTCCAAGAAGGTAGGTGC
MSS1950	GGCGAATTCTTGGGTCTTGGTTTCTCCTCTG
MSS1956	AGAATTGTAATACGACTCACTATAGGGC
MSS2002	CCGCTCGAGCTATGTCTAAGTCCGAGTCTCCCAAGGAGCC
MSS2003	CCGGAATTCTTAGAACCTCCTGCCACTGCCATAGCTACTG
MSS2699	GGCGGATCCATGAACGGCACCCTGGACC
MSS2700	GGCCTCGAGGTAGGGCTTGCTGTCATTCTTCG
MSS2759	GGCGGATCCATGGCTGTTAGTGTCACACCAATT
MSS2760	GGCCTCGAGCTGGTATTGGGCAGCCTTGA
MSS2129	GCATCTTCATGGTGTGGACATGCAAGAAAAAAGC
MSS2130	GCTTTTTTCTTGCATGTCCACACCATGAAGATGC
MSS3045	GCTGGCTAGCACCATGGGTAAGCCTATCC
MSS3046	GGCGGATCCCGTAGAATCGAGACCGAGGAGAGGGTTAGG
MSS3073	GGGCTCGAGCTATGGCTGTTAGTGTCACACCAATT
MSS3074	GGCGGATCCCTACTGGTATTGGGCAGCCTTGA
MSS3105	TAATACGACTCACTATAGGGCTTCTGCAGCTGCGTGGCTTTTTTC
MSS3106	CCTCGGCGACAGCATCTTCATGG
MSS3130	TAATACGACTCACTATAGGCCACTGCTGCTGTGTGGCC
MSS3145	GGCGGATCCATGCCCCCTGCACATTTACAAGCC
MSS3146	CGCCTCGAGCATCTGGGTAACATACTTGTGGCTAGTC
MSS3163	GGCTTTTTTCTTGCATGTGCACTTGTGTCCACACCATGAAGATGCTG
MSS3164	CAGCATCTTCATGGTGTGGACACAAGTGCACATGCAAGAAAAAAGCC
MSS3165	GGCTTTTTTCTTGCATGTGCGTTTGTACCCACACCATGAAGATGCTG
MSS3166	CAGCATCTTCATGGTGTGGGTACAAACGCACATGCAAGAAAAAAGCC
MSS3169	GGCTTTTTTCTTGCATGTGGGCTTGTGCCCACACCATGA
MSS3170	TCATGGTGTGGGCACAAGCCCACATGCAAGAAAAAAGCC

LIST OF REFERENCES

1. Harper, P.S. (2001) Myotonic Dystrophy.

2. Ranum, L.P. and Day, J.W. (2004) Myotonic dystrophy: RNA pathogenesis comes into focus. *Am. J. Hum. Genet.*, **74**, 793-804.

3. Liquori, C.L., Ricker, K., Moseley, M.L., Jacobsen, J.F., Kress, W., Naylor, S.L., Day, J.W. and Ranum, L.P. (2001) Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science*, **293**, 864-867.

4. Harley, H.G., Brook, J.D., Rundle, S.A., Crow, S., Reardon, W., Buckler, A.J., Harper, P.S., Housman, D.E. and Shaw, D.J. (1992) Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. *Nature*, **355**, 545-546.

5. Buxton, J., Shelbourne, P., Davies, J., Jones, C., Van Tongeren, T., Aslanidis, C., de Jong, P., Jansen, G., Anvret, M., Riley, B. *et al.* (1992) Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. *Nature*, **355**, 547-548.

6. Brook, J.D., McCurrach, M.E., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H., Hunter, K., Stanton, V.P., Thirion, J.P., Hudson, T. *et al.* (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell*, **68**, 799-808.

7. Aslanidis, C., Jansen, G., Amemiya, C., Shutler, G., Mahadevan, M., Tsilfidis, C., Chen, C., Alleman, J., Wormskamp, N.G., Vooijs, M. *et al.* (1992) Cloning of the essential myotonic dystrophy region and mapping of the putative defect. *Nature*, **355**, 548-551.

8. Larkin, K. and Fardaei, M. (2001) Myotonic dystrophy--a multigene disorder. *Brain Res. Bull.*, **56**, 389-395.

9. Machuca-Tzili, L., Brook, D. and Hilton-Jones, D. (2005) Clinical and molecular aspects of the myotonic dystrophies: a review. *Muscle Nerve*, **32**, 1-18.

10. Day, J.W. and Ranum, L.P. (2005) RNA pathogenesis of the myotonic dystrophies. *Neuromuscul. Disord.*, **15**, 5-16.

11. Cho, D.H. and Tapscott, S.J. (2007) Myotoinc dystrophy: Emerging mechanisms for DM1 and DM2. *Biochimica et Biophysica Acta*, 195-204.

12. Day, J.W., Ricker, K., Jacobsen, J.F., Rasmussen, L.J., Dick, K.A., Kress, W., Schneider, C., Koch, M.C., Beilman, G.J., Harrison, A.R. *et al.* (2003) Myotonic dystrophy type 2: molecular, diagnostic and clinical spectrum. *Neurology*, **60**, 657-664.

13. Fu, Y.H., Pizzuti, A., Fenwick, R.G., Jr., King, J., Rajnarayan, S., Dunne, P.W., Dubel, J., Nasser, G.A., Ashizawa, T., de Jong, P. *et al.* (1992) An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science*, **255**, 1256-1258.

14. Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barcelo, J., O'Hoy, K. *et al.* (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science*, **255**, 1253-1255.

15. Harley, H.G., Rundle, S.A., Reardon, W., Myring, J., Crow, S., Brook, J.D., Harper, P.S. and Shaw, D.J. (1992) Unstable DNA sequence in myotonic dystrophy. *Lancet*, **339**, 1125-1128.

16. Ashizawa, T., Dubel, J.R. and Harati, Y. (1993) Somatic instability of CTG repeat in myotonic dystrophy. *Neurology*, **43**, 2674-2678.

17. Martorell, L., Johnson, K., Boucher, C.A. and Baiget, M. (1997) Somatic instability of the myotonic dystrophy (CTG)n repeat during human fetal development. *Hum. Mol. Genet.*, **6**, 877-880.

18. Wong, L.J. and Ashizawa, T. (1997) Instability of the (CTG)n repeat in congenital myotonic dystrophy. *Am. J. Hum. Genet.*, **61**, 1445-1448.

19. Wong, L.J., Ashizawa, T., Monckton, D.G., Caskey, C.T. and Richards, C.S. (1995) Somatic heterogeneity of the CTG repeat in myotonic dystrophy is age and size dependent. *Am. J. Hum. Genet.*, **56**, 114-122.

20. Martorell, L., Monckton, D.G., Gamez, J., Johnson, K.J., Gich, I., Lopez de Munain, A. and Baiget, M. (1998) Progression of somatic CTG repeat length heterogeneity in the blood cells of myotonic dystrophy patients. *Hum. Mol. Genet.*, **7**, 307-312.

21. Abbruzzese, C., Costanzi Porrini, S., Mariani, B., Gould, F.K., McAbney, J.P., Monckton, D.G., Ashizawa, T. and Giacanelli, M. (2002) Instability of a premutation allele in homozygous patients with myotonic dystrophy type 1. *Annals of Neurology*, **52**, 435-441.

22. Duthel, S., Bost, M., Ollagnon, E., Vial, C., Petiot, P., Chazot, G. and Vandenberghe, A. (1999) CTG instability in myotonic dystrophy: molecular genetic analysis of families from south-eastern France with characteristics of intergenerational variation in CGT repeat numbers. *Annales de Genetique*, **42**, 151-159.

23. Brunner, H.G., Bruggenwirth, H.T., Nillesen, W., Jansen, G., Hamel, B.C., Hoppe, R.L., de Die, C.E., Howeler, C.J., van Oost, B.A., Wieringa, B. *et al.* (1993) Influence of sex of the transmitting parent as well as of parental allele size on the CTG expansion in myotonic dystrophy (DM). *Am. J. Hum. Genet.*, **53**, 1016-1023.

24. Lahue, R.S. and Slater, D.L. (2003) DNA repair and trinucleotide repeat instability. *Front. Biosci.*, **8**, s653-665.

25. Parniewski, P. and Staczek, P. (2002) Molecular mechanisms of TRS instability. *Advances in Experimental Medicine and Biology*, **516**, 1-25.

26. Cleary, J.D. and Pearson, C.E. (2003) The contribution of cis-elements to disease-associated repeat instability: clinical and experimental evidence. *Cytogenet. Genome Res.*, **100**, 25-55.

27. van den Broek, W.J., Nelen, M.R., Wansink, D.G., Coerwinkel, M.M., te Riele, H., Groenen, P.J. and Wieringa, B. (2002) Somatic expansion behaviour of the (CTG)n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. *Hum. Mol. Genet.*, **11**, 191-198.

28. Foiry, L., Dong, L., Savouret, C., Hubert, L., te Riele, H., Junien, C. and Gourdon, G. (2006) Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice. *Human Genetics*, **119**, 520-526.

29. Savouret, C., Garcia-Cordier, C., Megret, J., te Riele, H., Junien, C. and Gourdon, G. (2004) MSH2-dependent germinal CTG repeat expansions are produced continuously in spermatogonia from DM1 transgenic mice. *Mol. Cell. Biol.*, **24**, 629-637.

30. Manley, K., Shirley, T.L., Flaherty, L. and Messer, A. (1999) Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat. Genet.*, **23**, 471-473.

31. Kovtun, I.V. and McMurray, C.T. (2001) Trinucleotide expansion in haploid germ cells by gap repair. *Nat. Genet.*, **27**, 407-411.

32. Gomes-Pereira, M., Fortune, M.T., Ingram, L., McAbney, J.P. and Monckton, D.G. (2004) Pms2 is a genetic enhancer of trinucleotide CAG.CTG repeat somatic mosaicism: implications for the mechanism of triplet repeat expansion. *Hum. Mol. Genet.*, **13**, 1815-1825.

33. Stober, G., Franzek, E., Lesch, K.P. and Beckmann, H. (1995) Periodic catatonia: a schizophrenic subtype with major gene effect and anticipation. *European Archives of Psychiatry and Clinical Neuroscience*, **245**, 135-141.

34. Bassett, A.S. and Husted, J. (1997) Anticipation or ascertainment bias in schizophrenia? Penrose's familial mental illness sample. *Am. J. Hum. Genet.*, **60**, 630-637.

35. McInnis, M.G., McMahon, F.J., Chase, G.A., Simpson, S.G., Ross, C.A. and DePaulo, J.R., Jr. (1993) Anticipation in bipolar affective disorder. *Am. J. Hum. Genet.*, **53**, 385-390.

36. McInnis, M.G. (1996) Anticipation: an old idea in new genes. *Am. J. Hum. Genet.*, **59**, 973-979.

37. Ridley, R.M., Frith, C.D., Crow, T.J. and Conneally, P.M. (1988) Anticipation in Huntington's disease is inherited through the male line but may originate in the female. *Journal of Medical Genetics*, **25**, 589-595.

38. Manto, M.U. (2005) The wide spectrum of spinocerebellar ataxias (SCAs). *Cerebellum (London, England)*, **4**, 2-6.

39. Flink, I.L. and Morkin, E. (1995) Alternatively processed isoforms of cellular nucleic acid-binding protein interact with a suppressor region of the human beta-myosin heavy chain gene. *J. Biol. Chem.*, **270**, 6959-6965.

40. Gerbasi, V.R. and Link, A.J. (2007) The myotonic dystrophy type 2 protein ZNF9 is part of an ITAF complex that promotes cap-independent translation. *Mol. Cell Proteomics*, **6**, 1049-1058.

41. Margolis, J.M., Schoser, B.G., Moseley, M.L., Day, J.W. and Ranum, L.P. (2006) DM2 intronic expansions: evidence for CCUG accumulation without flanking sequence or effects on ZNF9 mRNA processing or protein expression. *Hum. Mol. Genet.*, **15**, 1808-1815.

42. Liquori, C.L., Ikeda, Y., Weatherspoon, M., Ricker, K., Schoser, B.G., Dalton, J.C., Day, J.W. and Ranum, L.P. (2003) Myotonic dystrophy type 2: human founder haplotype and evolutionary conservation of the repeat tract. *Am. J. Hum. Genet.*, **73**, 849-862.

43. Schoser, B.G., Kress, W., Walter, M.C., Halliger-Keller, B., Lochmuller, H. and Ricker, K. (2004) Homozygosity for CCTG mutation in myotonic dystrophy type 2. *Brain*, **127**, 1868-1877.

44. Fu, Y.H., Friedman, D.L., Richards, S., Pearlman, J.A., Gibbs, R.A., Pizzuti, A., Ashizawa, T., Perryman, M.B., Scarlato, G., Fenwick, R.G., Jr. *et al.* (1993) Decreased expression of myotonin-protein kinase messenger RNA and protein in adult form of myotonic dystrophy. *Science*, **260**, 235-238.

45. Novelli, G., Gennarelli, M., Zelano, G., Pizzuti, A., Fattorini, C., Caskey, C.T. and Dallapiccola, B. (1993) Failure in detecting mRNA transcripts from the mutated allele in myotonic dystrophy muscle. *Biochemistry and Molecular Biology International*, **29**, 291-297.

46. Hofmann-Radvanyi, H., Lavedan, C., Rabes, J.P., Savoy, D., Duros, C., Johnson, K. and Junien, C. (1993) Myotonic dystrophy: absence of CTG enlarged transcript in congenital forms, and low expression of the normal allele. *Hum. Mol. Genet.*, **2**, 1263-1266.

47. Narang, M.A., Waring, J.D., Sabourin, L.A. and Korneluk, R.G. (2000) Myotonic dystrophy (DM) protein kinase levels in congenital and adult DM patients. *Eur. J. Hum. Genet.*, **8**, 507-512.

48. Jansen, G., Groenen, P.J., Bachner, D., Jap, P.H., Coerwinkel, M., Oerlemans, F., van den Broek, W., Gohlsch, B., Pette, D., Plomp, J.J. *et al.* (1996) Abnormal myotonic dystrophy protein kinase levels produce only mild myopathy in mice. *Nat. Genet.*, **13**, 316-324.

49. Reddy, S., Smith, D.B., Rich, M.M., Leferovich, J.M., Reilly, P., Davis, B.M., Tran, K., Rayburn, H., Bronson, R., Cros, D. *et al.* (1996) Mice lacking the myotonic dystrophy protein kinase develop a late onset progressive myopathy. *Nat. Genet.*, **13**, 325-335.

50. Berul, C.I., Maguire, C.T., Gehrmann, J. and Reddy, S. (2000) Progressive atrioventricular conduction block in a mouse myotonic dystrophy model. *J. Interv. Card. Electrophysiol.*, **4**, 351-358.

51. Wang, Y.H., Amirhaeri, S., Kang, S., Wells, R.D. and Griffith, J.D. (1994) Preferential nucleosome assembly at DNA triplet repeats from the myotonic dystrophy gene. *Science*, **265**, 669-671.

52. Otten, A.D. and Tapscott, S.J. (1995) Triplet repeat expansion in myotonic dystrophy alters the adjacent chromatin structure. *Proc. Natl. Acad. Sci., U S A*, **92**, 5465-5469.

53. Klesert, T.R., Otten, A.D., Bird, T.D. and Tapscott, S.J. (1997) Trinucleotide repeat expansion at the myotonic dystrophy locus reduces expression of DMAHP. *Nat. Genet.*, **16**, 402-406.

54. Hamshere, M.G., Newman, E.E., Alwazzan, M., Athwal, B.S. and Brook, J.D. (1997) Transcriptional abnormality in myotonic dystrophy affects DMPK but not neighboring genes. *Proc. Natl. Acad. Sci., U S A*, **94**, 7394-7399.

55. Frisch, R., Singleton, K.R., Moses, P.A., Gonzalez, I.L., Carango, P., Marks, H.G. and Funanage, V.L. (2001) Effect of triplet repeat expansion on chromatin structure and expression of DMPK and neighboring genes, SIX5 and DMWD, in myotonic dystrophy. *Molecular Genetics and Metabolism*, **74**, 281-291.

56. Thornton, C.A., Wymer, J.P., Simmons, Z., McClain, C. and Moxley, R.T., 3rd. (1997) Expansion of the myotonic dystrophy CTG repeat reduces expression of the flanking DMAHP gene. *Nat. Genet.*, **16**, 407-409.

57. Klesert, T.R., Cho, D.H., Clark, J.I., Maylie, J., Adelman, J., Snider, L., Yuen, E.C., Soriano, P. and Tapscott, S.J. (2000) Mice deficient in Six5 develop cataracts: implications for myotonic dystrophy. *Nat. Genet.*, **25**, 105-109.

58. Sarkar, P.S., Appukuttan, B., Han, J., Ito, Y., Ai, C., Tsai, W., Chai, Y., Stout, J.T. and Reddy, S. (2000) Heterozygous loss of Six5 in mice is sufficient to cause ocular cataracts. *Nat. Genet.*, **25**, 110-114.

59. Taneja, K.L., McCurrach, M., Schalling, M., Housman, D. and Singer, R.H. (1995) Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J. Cell Biol.*, **128**, 995-1002.

60. Timchenko, L.T., Miller, J.W., Timchenko, N.A., DeVore, D.R., Datar, K.V., Lin, L., Roberts, R., Caskey, C.T. and Swanson, M.S. (1996) Identification of a (CUG)n triplet repeat RNA-binding protein and its expression in myotonic dystrophy. *Nucleic Acids Res.*, **24**, 4407-4414.

61. Mankodi, A., Logigian, E., Callahan, L., McClain, C., White, R., Henderson, D., Krym, M. and Thornton, C.A. (2000) Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science*, **289**, 1769-1773.

62. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406-3415.

63. Michalowski, S., Miller, J.W., Urbinati, C.R., Paliouras, M., Swanson, M.S. and Griffith, J. (1999) Visualization of double-stranded RNAs from the myotonic dystrophy protein kinase gene and interactions with CUG-binding protein. *Nucleic Acids Res.*, **27**, 3534-3542.

64. Sobczak, K., de Mezer, M., Michlewski, G., Krol, J. and Krzyzosiak, W.J. (2003) RNA structure of trinucleotide repeats associated with human neurological diseases. *Nucleic Acids Res.*, **31**, 5469-5482.

65. Mooers, B.H., Logue, J.S. and 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**, 16626-16631.

66. Tian, B., White, R.J., Xia, T., Welle, S., Turner, D.H., Mathews, M.B. and Thornton, C.A. (2000) Expanded CUG repeat RNAs form hairpins that activate the double-stranded RNA-dependent protein kinase PKR. *RNA (New York, N.Y)*, **6**, 79-87.

67. Napierala, M. and Krzyzosiak, W.J. (1997) CUG repeats present in myotonin kinase RNA form metastable "slippery" hairpins. *J. Biol. Chem.*, **272**, 31079-31085.

68. 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. *The EMBO Journal*, **19**, 4439-4448.

69. Kanadia, R.N., Johnstone, K.A., Mankodi, A., Lungu, C., Thornton, C.A., Esson, D., Timmers, A.M., Hauswirth, W.W. and Swanson, M.S. (2003) A muscleblind knockout model for myotonic dystrophy. *Science*, **302**, 1978-1980.

70. Ebralidze, A., Wang, Y., Petkova, V., Ebralidse, K. and Junghans, R.P. (2004) RNA leaching of transcription factors disrupts transcription in myotonic dystrophy. *Science*, **303**, 383-387.

71. Artero, R., Prokop, A., Paricio, N., Begemann, G., Pueyo, I., Mlodzik, M., Perez-Alonso, M. and Baylies, M.K. (1998) The muscleblind gene participates in the organization of Z-bands and epidermal attachments of Drosophila muscles and is regulated by Dmef2. *Dev. Biol.*, **195**, 131-143.

72. 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.

73. Pascual, M., Vicente, M., Monferrer, L. and Artero, R. (2006) The Muscleblind family of proteins: an emerging class of regulators of developmentally programmed alternative splicing. *Differentiation*, **74**, 65-80.

74. Gomperts, M., Pascall, J.C. and Brown, K.D. (1990) The nucleotide sequence of a cDNA encoding an EGF-inducible gene indicates the existence of a new family of mitogen-induced genes. *Oncogene*, **5**, 1081-1083.

75. Anderson, J.T., Wilson, S.M., Datar, K.V. and Swanson, M.S. (1993) NAB2: a yeast nuclear polyadenylated RNA-binding protein essential for cell viability. *Mol. Cell. Biol.*, **13**, 2730-2741.

76. Hudson, B.P., Martinez-Yamout, M.A., Dyson, H.J. and Wright, P.E. (2004) Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. *Nat. Struct. Mol. Biol.*, **11**, 257-264.

77. Monferrer, L. and Artero, R. (2006) An interspecific functional complementation test in Drosophila for introductory genetics laboratory courses. *The Journal of Heredity*, **97**, 67-73.

78. Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S. and Haussler, D. (2004) Ultraconserved elements in the human genome. *Science*, **304**, 1321-1325.

79. Fardaei, M., Rogers, M.T., Thorpe, H.M., Larkin, K., Hamshere, M.G., Harper, P.S. and Brook, J.D. (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. Genet.*, **11**, 805-814.

80. Kanadia, R.N., Urbinati, C.R., Crusselle, V.J., Luo, D., Lee, Y.J., Harrison, J.K., Oh, S.P. and Swanson, M.S. (2003) Developmental expression of mouse muscleblind genes Mbnl1, Mbnl2 and Mbnl3. *Gene Expr. Patterns*, **3**, 459-462.

81. Wang, G.S. and Cooper, T.A. (2007) Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat. Rev. Genet.*, **8**, 749-761.

82. Stamm, S., Zhu, J., Nakai, K., Stoilov, P., Stoss, O. and Zhang, M.Q. (2000) An alternative-exon database and its statistical analysis. *DNA and Cell Biology*, **19**, 739-756.

83. Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. and Sharp, P.A. (1986) Splicing of messenger RNA precursors. *Annual Review of Biochemistry*, **55**, 1119-1150.

84. Staley, J.P. and Guthrie, C. (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell*, **92**, 315-326.

85. Wu, S., Romfo, C.M., Nilsen, T.W. and Green, M.R. (1999) Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature*, **402**, 832-835.

86. Stamm, S., Zhang, M.Q., Marr, T.G. and Helfman, D.M. (1994) A sequence compilation and comparison of exons that are alternatively spliced in neurons. *Nucleic Acids Res.*, **22**, 1515-1526.

87. Clark, F. and Thanaraj, T.A. (2002) Categorization and characterization of transcript-confirmed constitutively and alternatively spliced introns and exons from human. *Hum. Mol. Genet.*, **11**, 451-464.

88. Kanopka, A., Muhlemann, O. and Akusjarvi, G. (1996) Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature*, **381**, 535-538.

89. Cooper, T.A. and Mattox, W. (1997) The regulation of splice-site selection, and its role in human disease. *Am. J. Hum. Genet.*, **61**, 259-266.

90. Damgaard, C.K., Tange, T.O. and Kjems, J. (2002) hnRNP A1 controls HIV-1 mRNA splicing through cooperative binding to intron and exon splicing silencers in the context of a conserved secondary structure. *RNA (New York, N.Y)*, **8**, 1401-1415.

91. Tange, T.O., Damgaard, C.K., Guth, S., Valcarcel, J. and Kjems, J. (2001) The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. *The EMBO Journal*, **20**, 5748-5758.

92. Carstens, R.P., Wagner, E.J. and Garcia-Blanco, M.A. (2000) An intronic splicing silencer causes skipping of the IIIb exon of fibroblast growth factor receptor 2 through involvement of polypyrimidine tract binding protein. *Mol. Cell. Biol.*, **20**, 7388-7400.

93. Wagner, E.J. and Garcia-Blanco, M.A. (2001) Polypyrimidine tract binding protein antagonizes exon definition. *Mol. Cell. Biol.*, **21**, 3281-3288.

94. Lou, H., Helfman, D.M., Gagel, R.F. and Berget, S.M. (1999) Polypyrimidine tract-binding protein positively regulates inclusion of an alternative 3'-terminal exon. *Mol. Cell. Biol.*, **19**, 78-85.

95. Hofmann, Y. and Wirth, B. (2002) hnRNP-G promotes exon 7 inclusion of survival motor neuron (SMN) via direct interaction with Htra2-beta1. *Hum. Mol. Genet.*, **11**, 2037-2049.

96. Chou, M.Y., Rooke, N., Turck, C.W. and Black, D.L. (1999) hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. *Mol. Cell. Biol.*, **19**, 69-77.

97. Min, H., Chan, R.C. and Black, D.L. (1995) The generally expressed hnRNP F is involved in a neural-specific pre-mRNA splicing event. *Genes & Development*, **9**, 2659-2671.

98. Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B.J. and Darnell, R.B. (2006) An RNA map predicting Nova-dependent splicing regulation. *Nature*, **444**, 580-586.

99. Zhu, J., Mayeda, A. and Krainer, A.R. (2001) Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol. Cell*, **8**, 1351-1361.

100. Ashiya, M. and Grabowski, P.J. (1997) A neuron-specific splicing switch mediated by an array of pre-mRNA repressor sites: evidence of a regulatory role for the polypyrimidine tract binding protein and a brain-specific PTB counterpart. *RNA (New York, N.Y)*, **3**, 996-1015.

101. Irwin, N., Baekelandt, V., Goritchenko, L. and Benowitz, L.I. (1997) Identification of two proteins that bind to a pyrimidine-rich sequence in the 3'-untranslated region of GAP-43 mRNA. *Nucleic Acids Res.*, **25**, 1281-1288.

102. Markovtsov, V., Nikolic, J.M., Goldman, J.A., Turck, C.W., Chou, M.Y. and Black, D.L. (2000) Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. *Mol. Cell. Biol.*, **20**, 7463-7479.

103. Coutinho-Mansfield, G.C., Xue, Y., Zhang, Y. and Fu, X.D. (2007) PTB/nPTB switch: a post-transcriptional mechanism for programming neuronal differentiation. *Genes & Development*, **21**, 1573-1577.

104. Boutz, P.L., Stoilov, P., Li, Q., Lin, C.H., Chawla, G., Ostrow, K., Shiue, L., Ares, M., Jr. and Black, D.L. (2007) A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes & Development*, **21**, 1636-1652.

105. Shin, C., Feng, Y. and Manley, J.L. (2004) Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock. *Nature*, **427**, 553-558.

106. Shin, C. and Manley, J.L. (2004) Cell signalling and the control of pre-mRNA splicing. *Nature Reviews*, **5**, 727-738.

107. Krawczak, M., Reiss, J. and Cooper, D.N. (1992) The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Human Genetics*, **90**, 41-54.

108. Pagani, F. and Baralle, F.E. (2004) Genomic variants in exons and introns: identifying the splicing spoilers. *Nat. Rev. Genet.*, **5**, 389-396.

109. Lopez-Bigas, N., Audit, B., Ouzounis, C., Parra, G. and Guigo, R. (2005) Are splicing mutations the most frequent cause of hereditary disease? *FEBS Letters*, **579**, 1900-1903.

110. Osborne, R.J. and Thornton, C.A. (2006) RNA-dominant diseases. *Hum. Mol. Genet.*, **15 Spec No 2**, R162-169.

111. Mankodi, A., Takahashi, M.P., Jiang, H., Beck, C.L., Bowers, W.J., Moxley, R.T., Cannon, S.C. and Thornton, C.A. (2002) Expanded CUG repeats trigger aberrant splicing of ClC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy. *Mol. Cell*, **10**, 35-44.

112. Charlet, B.N., Savkur, R.S., Singh, G., Philips, A.V., Grice, E.A. and Cooper, T.A. (2002) Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. *Mol. Cell*, **10**, 45-53.

113. Savkur, R.S., Philips, A.V. and Cooper, T.A. (2001) Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. *Nat. Genet.*, **29**, 40-47.

114. Moxley, R.T., Griggs, R.C. and Goldblatt, D. (1980) Muscle insulin resistance in myotonic dystrophy: effect of supraphysiologic insulinization. *Neurology*, **30**, 1077-1083.

115. 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.

116. Kino, Y., Mori, D., Oma, Y., Takeshita, Y., Sasagawa, N. and Ishiura, S. (2004) Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. *Hum. Mol. Genet.*, **13**, 495-507.

117. Rooke, N., Markovtsov, V., Cagavi, E. and Black, D.L. (2003) Roles for SR proteins and hnRNP A1 in the regulation of c-src exon N1. *Mol. Cell. Biol.*, **23**, 1874-1884.

118. Ho, T.H., Charlet, B.N., Poulos, M.G., Singh, G., Swanson, M.S. and Cooper, T.A. (2004) Muscleblind proteins regulate alternative splicing. *The EMBO Journal*, **23**, 3103-3112.

119. Vicente, M., Monferrer, L., Poulos, M.G., Houseley, J., Monckton, D.G., O'Dell K, M., Swanson, M.S. and Artero, R.D. (2007) Muscleblind isoforms are functionally distinct and regulate alpha-actinin splicing. *Differentiation*, **75**, 427-440.

120. Hino, S.I., Kondo, S., Sekiya, H., Saito, A., Kanemoto, S., Murakami, T., Chihara, K., Aoki, Y., Nakamori, M., Takahashi, M.P. *et al.* (2007) Molecular mechanisms responsible for aberrant splicing of SERCA1 in myotonic dystrophy type 1. *Hum. Mol. Genet.*, **16**, 2834-2843.

121. Sharma, S. and Black, D.L. (2006) Maps, codes, and sequence elements: can we predict the protein output from an alternatively spliced locus? *Neuron*, **52**, 574-576.

122. Lin, X., Miller, J.W., Mankodi, A., Kanadia, R.N., Yuan, Y., Moxley, R.T., Swanson, M.S. and Thornton, C.A. (2006) Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy. *Hum. Mol. Genet.*, **15**, 2087-2097.

123. Makeyev, E.V., Zhang, J., Carrasco, M.A. and Maniatis, T. (2007) The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol. Cell*, **27**, 435-448.

124. Ule, J., Jensen, K., Mele, A. and Darnell, R.B. (2005) CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods (San Diego, Calif.)*, **37**, 376-386.

125. Timchenko, N.A., Cai, Z.J., Welm, A.L., Reddy, S., Ashizawa, T. and Timchenko, L.T. (2001) RNA CUG repeats sequester CUGBP1 and alter protein levels and activity of CUGBP1. *J. Biol. Chem.*, **276**, 7820-7826.

126. Savkur, R.S., Philips, A.V., Cooper, T.A., Dalton, J.C., Moseley, M.L., Ranum, L.P. and Day, J.W. (2004) Insulin receptor splicing alteration in myotonic dystrophy type 2. *Am. J. Hum. Genet.*, **74**, 1309-1313.

127. Dansithong, W., Paul, S., Comai, L. and Reddy, S. (2005) MBNL1 is the primary determinant of focus formation and aberrant insulin receptor splicing in DM1. *J. Biol. Chem.*, **280**, 5773-5780.

128. Lefebvre, S., Burlet, P., Viollet, L., Bertrandy, S., Huber, C., Belser, C. and Munnich, A. (2002) A novel association of the SMN protein with two major non-ribosomal nucleolar proteins and its implication in spinal muscular atrophy. *Hum. Mol. Genet.*, **11**, 1017-1027.

129. Srivastava, S., Mukherjee, M., Panigrahi, I., Shanker Pandey, G., Pradhan, S. and Mittal, B. (2001) SMN2-deletion in childhood-onset spinal muscular atrophy. *American Journal of Medical Genetics*, **101**, 198-202.

130. Underwood, J.G., Boutz, P.L., Dougherty, J.D., Stoilov, P. and Black, D.L. (2005) Homologues of the Caenorhabditis elegans Fox-1 protein are neuronal splicing regulators in mammals. *Mol. Cell. Biol.*, **25**, 10005-10016.

131. Mutsuddi, M., Marshall, C.M., Benzow, K.A., Koob, M.D. and Rebay, I. (2004) The spinocerebellar ataxia 8 noncoding RNA causes neurodegeneration and associates with staufen in Drosophila. *Curr. Biol.*, **14**, 302-308.

132. Wakamiya, M., Matsuura, T., Liu, Y., Schuster, G.C., Gao, R., Xu, W., Sarkar, P.S., Lin, X. and Ashizawa, T. (2006) The role of ataxin 10 in the pathogenesis of spinocerebellar ataxia type 10. *Neurology*, **67**, 607-613.

133. Rudnicki, D.D., Holmes, S.E., Lin, M.W., Thornton, C.A., Ross, C.A. and Margolis, R.L. (2007) Huntington's disease--like 2 is associated with CUG repeat-containing RNA foci. *Annals of Neurology*, **61**, 272-282.

134. Iwahashi, C.K., Yasui, D.H., An, H.J., Greco, C.M., Tassone, F., Nannen, K., Babineau, B., Lebrilla, C.B., Hagerman, R.J. and Hagerman, P.J. (2006) Protein composition of the intranuclear inclusions of FXTAS. *Brain*, **129**, 256-271.

135. Gabellini, D., D'Antona, G., Moggio, M., Prelle, A., Zecca, C., Adami, R., Angeletti, B., Ciscato, P., Pellegrino, M.A., Bottinelli, R. *et al.* (2006) Facioscapulohumeral muscular dystrophy in mice overexpressing FRG1. *Nature*, **439**, 973-977.

136. Ule, J., Ule, A., Spencer, J., Williams, A., Hu, J.S., Cline, M., Wang, H., Clark, T., Fraser, C., Ruggiu, M. *et al.* (2005) Nova regulates brain-specific splicing to shape the synapse. *Nat. Genet.*, **37**, 844-852.

137. Jensen, K.B., Dredge, B.K., Stefani, G., Zhong, R., Buckanovich, R.J., Okano, H.J., Yang, Y.Y. and Darnell, R.B. (2000) Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron*, **25**, 359-371.

138. Bhalla, K., Phillips, H.A., Crawford, J., McKenzie, O.L., Mulley, J.C., Eyre, H., Gardner, A.E., Kremmidiotis, G. and Callen, D.F. (2004) The de novo chromosome 16 translocations of two patients with abnormal phenotypes (mental retardation and epilepsy) disrupt the A2BP1 gene. *Journal of Human Genetics*, **49**, 308-311.

139. McKie, A.B., McHale, J.C., Keen, T.J., Tarttelin, E.E., Goliath, R., van Lith-Verhoeven, J.J., Greenberg, J., Ramesar, R.S., Hoyng, C.B., Cremers, F.P. *et al.* (2001) Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13). *Hum. Mol. Genet.*, **10**, 1555-1562.

140. Vithana, E.N., Abu-Safieh, L., Allen, M.J., Carey, A., Papaioannou, M., Chakarova, C., Al-Maghtheh, M., Ebenezer, N.D., Willis, C., Moore, A.T. *et al.* (2001) A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). *Mol. Cell*, **8**, 375-381.

141. Chakarova, C.F., Hims, M.M., Bolz, H., Abu-Safieh, L., Patel, R.J., Papaioannou, M.G., Inglehearn, C.F., Keen, T.J., Willis, C., Moore, A.T. *et al.* (2002) Mutations in HPRP3, a third member of pre-mRNA splicing factor genes, implicated in autosomal dominant retinitis pigmentosa. *Hum. Mol. Genet.*, **11**, 87-92.

142. Yang, L., Embree, L.J. and Hickstein, D.D. (2000) TLS-ERG leukemia fusion protein inhibits RNA splicing mediated by serine-arginine proteins. *Mol. Cell. Biol.*, **20**, 3345-3354.

143. Ichikawa, H., Shimizu, K., Hayashi, Y. and Ohki, M. (1994) An RNA-binding protein gene, TLS/FUS, is fused to ERG in human myeloid leukemia with t(16;21) chromosomal translocation. *Cancer Research*, **54**, 2865-2868.

144. Venables, J.P., Elliott, D.J., Makarova, O.V., Makarov, E.M., Cooke, H.J. and Eperon, I.C. (2000) RBMY, a probable human spermatogenesis factor, and other hnRNP G proteins interact with Tra2beta and affect splicing. *Hum. Mol. Genet.*, **9**, 685-694.

145. Imai, H., Chan, E.K., Kiyosawa, K., Fu, X.D. and Tan, E.M. (1993) Novel nuclear autoantigen with splicing factor motifs identified with antibody from hepatocellular carcinoma. *J. Clin. Invest.*, **92**, 2419-2426.

146. Clark, J., Lu, Y.J., Sidhar, S.K., Parker, C., Gill, S., Smedley, D., Hamoudi, R., Linehan, W.M., Shipley, J. and Cooper, C.S. (1997) Fusion of splicing factor genes PSF and NonO (p54nrb) to the TFE3 gene in papillary renal cell carcinoma. *Oncogene*, **15**, 2233-2239.

147. Young, J.I., Hong, E.P., Castle, J.C., Crespo-Barreto, J., Bowman, A.B., Rose, M.F., Kang, D., Richman, R., Johnson, J.M., Berget, S. *et al.* (2005) Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *Proc. Natl. Acad. Sci., U S A*, **102**, 17551-17558.

148. Kishore, S. and Stamm, S. (2006) The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science*, **311**, 230-232.

149. Fomenkov, A., Huang, Y.P., Topaloglu, O., Brechman, A., Osada, M., Fomenkova, T., Yuriditsky, E., Trink, B., Sidransky, D. and Ratovitski, E. (2003) P63 alpha mutations lead to aberrant splicing of keratinocyte growth factor receptor in the Hay-Wells syndrome. *J. Biol. Chem.*, **278**, 23906-23914.

150. Ho, T.H., Savkur, R.S., Poulos, M.G., Mancini, M.A., Swanson, M.S. and Cooper, T.A. (2005) Colocalization of muscleblind with RNA foci is separable from mis-regulation of alternative splicing in myotonic dystrophy. *Journal of Cell Science*, **118**, 2923-2933.

151. Blackshear, P.J. (2002) Tristetraprolin and other CCCH tandem zinc-finger proteins in the regulation of mRNA turnover. *Biochemical Society Transactions*, **30**, 945-952.

152. Nykamp, K.R. and Swanson, M.S. (2004) Toxic RNA in the nucleus: unstable microsatellite expression in neuromuscular disease. *Progress in Molecular and Subcellular Biology*, **35**, 57-77.

153. Mankodi, A., Teng-Umnuay, P., Krym, M., Henderson, D., Swanson, M. and Thornton, C.A. (2003) Ribonuclear inclusions in skeletal muscle in myotonic dystrophy types 1 and 2. *Annals of Neurology*, **54**, 760-768.

154. Mankodi, A., Lin, X., Blaxall, B.C., Swanson, M.S. and Thornton, C.A. (2005) Nuclear RNA foci in the heart in myotonic dystrophy. *Circulation Research*, **97**, 1152-1155.

155. Mankodi, A., Urbinati, C.R., Yuan, Q.P., Moxley, R.T., Sansone, V., Krym, M., Henderson, D., Schalling, M., Swanson, M.S. and Thornton, C.A. (2001) Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Hum. Mol. Genet.*, **10**, 2165-2170.

156. Jiang, H., Mankodi, A., Swanson, M.S., Moxley, R.T. and Thornton, C.A. (2004) Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Hum. Mol. Genet.*, **13**, 3079-3088.

157. Moore, M.J. (2005) From birth to death: the complex lives of eukaryotic mRNAs. *Science*, **309**, 1514-1518.

158. Cartegni, L., Maconi, M., Morandi, E., Cobianchi, F., Riva, S. and Biamonti, G. (1996) hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. *J. Mol. Biol.*, **259**, 337-348.

159. Kanadia, R.N., Shin, J., Yuan, Y., Beattie, S.G., Wheeler, T.M., Thornton, C.A. and Swanson, M.S. (2006) Reversal of RNA missplicing and myotonia after muscleblind overexpression in a mouse poly(CUG) model for myotonic dystrophy. *Proc. Natl. Acad. Sci., U S A*, **103**, 11748-11753.

160. Mahadevan, M.S., Yadava, R.S., Yu, Q., Balijepalli, S., Frenzel-McCardell, C.D., Bourne, T.D. and Phillips, L.H. (2006) Reversible model of RNA toxicity and cardiac conduction defects in myotonic dystrophy. *Nat. Genet.*, **38**, 1066-1070.

161. Morris, N.P., Oxford, J.T., Davies, G.B., Smoody, B.F. and Keene, D.R. (2000) Developmentally regulated alternative splicing of the alpha1(XI) collagen chain: spatial and temporal segregation of isoforms in the cartilage of fetal rat long bones. *J. Histochem. Cytochem.*, **48**, 725-741.

162. Thomas, W.S., O'Dowd, D.K. and Smith, M.A. (1993) Developmental expression and alternative splicing of chick agrin RNA. *Dev. Biol.*, **158**, 523-535.

163. Muro, A.F., Chauhan, A.K., Gajovic, S., Iaconcig, A., Porro, F., Stanta, G. and Baralle, F.E. (2003) Regulated splicing of the fibronectin EDA exon is essential for proper skin wound healing and normal lifespan. *J. Cell Biol.*, **162**, 149-160.

164. Filippova, G.N., Thienes, C.P., Penn, B.H., Cho, D.H., Hu, Y.J., Moore, J.M., Klesert, T.R., Lobanenkov, V.V. and Tapscott, S.J. (2001) CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. *Nat. Genet.*, **28**, 335-343.

165. Sabouri, L.A., Mahadevan, M.S., Narang, M., Lee, D.S., Surh, L.C. and Korneluk, R.G. (1993) Effect of the myotonic dystrophy (DM) mutation on mRNA levels of the DM gene. *Nat. Genet.*, **4**, 233-238.

166. Koob, M.D., Moseley, M.L., Schut, L.J., Benzow, K.A., Bird, T.D., Day, J.W. and Ranum, L.P. (1999) An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). *Nat. Genet.*, **21**, 379-384.

167. Day, J.W., Schut, L.J., Moseley, M.L., Durand, A.C. and Ranum, L.P. (2000) Spinocerebellar ataxia type 8: clinical features in a large family. *Neurology*, **55**, 649-657.

168. Topisirovic, I., Dragasevic, N., Savic, D., Ristic, A., Keckarevic, M., Keckarevic, D., Culjkovic, B., Petrovic, I., Romac, S. and Kostic, V.S. (2002) Genetic and clinical analysis of spinocerebellar ataxia type 8 repeat expansion in Yugoslavia. *Clinical Genetics*, **62**, 321-324.

169. Ikeda, Y., Shizuka, M., Watanabe, M., Okamoto, K. and Shoji, M. (2000) Molecular and clinical analyses of spinocerebellar ataxia type 8 in Japan. *Neurology*, **54**, 950-955.

170. Cellini, E., Nacmias, B., Forleo, P., Piacentini, S., Guarnieri, B.M., Serio, A., Calabro, A., Renzi, D. and Sorbi, S. (2001) Genetic and clinical analysis of spinocerebellar ataxia type 8 repeat expansion in Italy. *Archives of Neurology*, **58**, 1856-1859.

171. Stevanin, G., Herman, A., Durr, A., Jodice, C., Frontali, M., Agid, Y. and Brice, A. (2000) Are (CTG)n expansions at the SCA8 locus rare polymorphisms? *Nat. Genet.*, **24**, 213; author reply 215.

172. Moseley, M.L., Schut, L.J., Bird, T.D., Koob, M.D., Day, J.W. and Ranum, L.P. (2000) SCA8 CTG repeat: en masse contractions in sperm and intergenerational sequence changes may play a role in reduced penetrance. *Hum. Mol. Genet.*, **9**, 2125-2130.

173. Moseley, M.L., Zu, T., Ikeda, Y., Gao, W., Mosemiller, A.K., Daughters, R.S., Chen, G., Weatherspoon, M.R., Clark, H.B., Ebner, T.J. *et al.* (2006) Bidirectional expression of

CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. *Nat. Genet.*, **38**, 758-769.

174. Walker, R.H., Rasmussen, A., Rudnicki, D., Holmes, S.E., Alonso, E., Matsuura, T., Ashizawa, T., Davidoff-Feldman, B. and Margolis, R.L. (2003) Huntington's disease--like 2 can present as chorea-acanthocytosis. *Neurology*, **61**, 1002-1004.

175. Walker, R.H., Morgello, S., Davidoff-Feldman, B., Melnick, A., Walsh, M.J., Shashidharan, P. and Brin, M.F. (2002) Autosomal dominant chorea-acanthocytosis with polyglutamine-containing neuronal inclusions. *Neurology*, **58**, 1031-1037.

176. Walker, R.H., Jankovic, J., O'Hearn, E. and Margolis, R.L. (2003) Phenotypic features of Huntington's disease-like 2. *Mov. Disord.*, **18**, 1527-1530.

177. Stevanin, G., Camuzat, A., Holmes, S.E., Julien, C., Sahloul, R., Dode, C., Hahn-Barma, V., Ross, C.A., Margolis, R.L., Durr, A. *et al.* (2002) CAG/CTG repeat expansions at the Huntington's disease-like 2 locus are rare in Huntington's disease patients. *Neurology*, **58**, 965-967.

178. Holmes, S.E., O'Hearn, E., Rosenblatt, A., Callahan, C., Hwang, H.S., Ingersoll-Ashworth, R.G., Fleisher, A., Stevanin, G., Brice, A., Potter, N.T. *et al.* (2001) A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2. *Nat. Genet.*, **29**, 377-378.

179. Margolis, R.L., Holmes, S.E., Rosenblatt, A., Gourley, L., O'Hearn, E., Ross, C.A., Seltzer, W.K., Walker, R.H., Ashizawa, T., Rasmussen, A. *et al.* (2004) Huntington's Disease-like 2 (HDL2) in North America and Japan. *Annals of Neurology*, **56**, 670-674.

180. Margolis, R.L., O'Hearn, E., Rosenblatt, A., Willour, V., Holmes, S.E., Franz, M.L., Callahan, C., Hwang, H.S., Troncoso, J.C. and Ross, C.A. (2001) A disorder similar to Huntington's disease is associated with a novel CAG repeat expansion. *Annals of Neurology*, **50**, 373-380.

181. Berry-Kravis, E., Abrams, L., Coffey, S.M., Hall, D.A., Greco, C., Gane, L.W., Grigsby, J., Bourgeois, J.A., Finucane, B., Jacquemont, S. *et al.* (2007) Fragile X-associated tremor/ataxia syndrome: Clinical features, genetics, and testing guidelines. *Mov. Disord.*, **22**, 2018-2030.

182. Oostra, B.A. and Willemsen, R. (2003) A fragile balance: FMR1 expression levels. *Hum. Mol. Genet.*, **12 Spec No 2**, R249-257.

183. Hagerman, P.J. and Hagerman, R.J. (2004) The fragile-X premutation: a maturing perspective. *Am. J. Hum. Genet.*, **74**, 805-816.

184. Willemsen, R., Hoogeveen-Westerveld, M., Reis, S., Holstege, J., Severijnen, L.A., Nieuwenhuizen, I.M., Schrier, M., van Unen, L., Tassone, F., Hoogeveen, A.T. *et al.*

(2003) The FMR1 CGG repeat mouse displays ubiquitin-positive intranuclear neuronal inclusions; implications for the cerebellar tremor/ataxia syndrome. *Hum. Mol. Genet.*, **12**, 949-959.

185. Jin, P., Zarnescu, D.C., Zhang, F., Pearson, C.E., Lucchesi, J.C., Moses, K. and Warren, S.T. (2003) RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in Drosophila. *Neuron*, **39**, 739-747.

186. Warf, M.B. and 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 *(New York, N.Y)*, **13**, 2238-2251.

BIOGRAPHICAL SKETCH

Yuan Yuan was born in Wuhan, China on August 25th. She is the only child of Lizhi Yuan and Xinyun Zhu. Yuan attended Wuhan University from 1997 to 2001, during which time she pursued a Bachelor of Science degree in biological sciences. After graduation, Yuan came to University of Florida and joined the Interdisciplinary Program in Biomedical Sciences (IDP) in the College of Medicine. In May of 2002, Yuan chose to perform her doctorate research under the guidance of Dr. Maurice Swanson. She completed her graduate work and doctoral dissertation in December of 2007, and will move to Rockefeller University to begin her post-doctoral studies with Dr. Robert Darnell.