Supplemental Information

Muscleblind-Like 2 Mediated Alternative Splicing in the Developing Brain and Dysregulation in Myotonic Dystrophy

Konstantinos Charizanis, Kuang-Yung Lee, Ranjan Batra, Marianne Goodwin, Chaolin Zhang, Yuan Yuan, Lily Shiue, Melissa Cline, Marina M. Scotti, Guangbin Xia, Ashok Kumar, Tetsuo Ashizawa, H. Brent Clark, Takashi Kimura, Masanori P. Takahashi, Harutoshi Fujimura, Kenji Jinnai, Hiroo Yoshikawa, Mário Gomes-Pereira, Geneviève Gourdon, Noriaki Sakai, Seiji Nishino, Thomas C. Foster, Manuel Ares Jr, Robert B. Darnell and Maurice S. Swanson

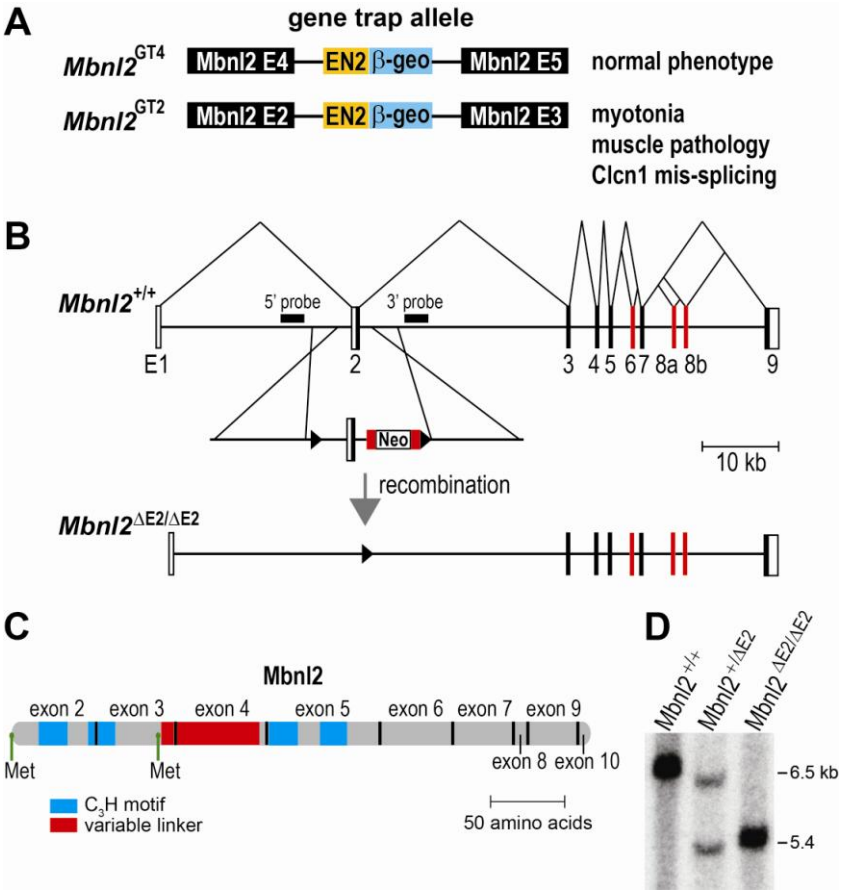

Figure S1. *Mbnl2* Gene Trap and Knockout Mice

(A) Gene trap alleles in which an engrailed 2 β-geo cassette has been inserted into either *Mbnl2* intron 4 (top) or intron 2 and the reported phenotypes (right).

(B) Homologous recombination of *Mbnl2*$^{\Delta E2}$ allele generation in ES cells. The Neo cassette (box), flanking loxP sites (triangles) and Frt sites (red regions flanking Neo) are indicated together with *Mbnl2* noncoding (open boxes) and coding (black and red boxes) regions. *Mbnl2* alternative cassette exons are also highlighted (red exons).

(C) Mbnl2 protein primary structure (grey rod). Shown are the two potential translation initiation sites (Met), the regions encompassing the CCCH (C3H) and variable linker motifs and the vertical black lines separate the regions encoded by *Mbnl2* exons 2-9.

(D) Genomic DNA blot of *Stu*I digested *Mbnl2*$^{+/+}$, *Mbnl2*$^{+/\Delta E2}$ and *Mbnl2*$^{\Delta E2/\Delta E2}$ DNAs hybridized with the 5' probe (Figure S1B).

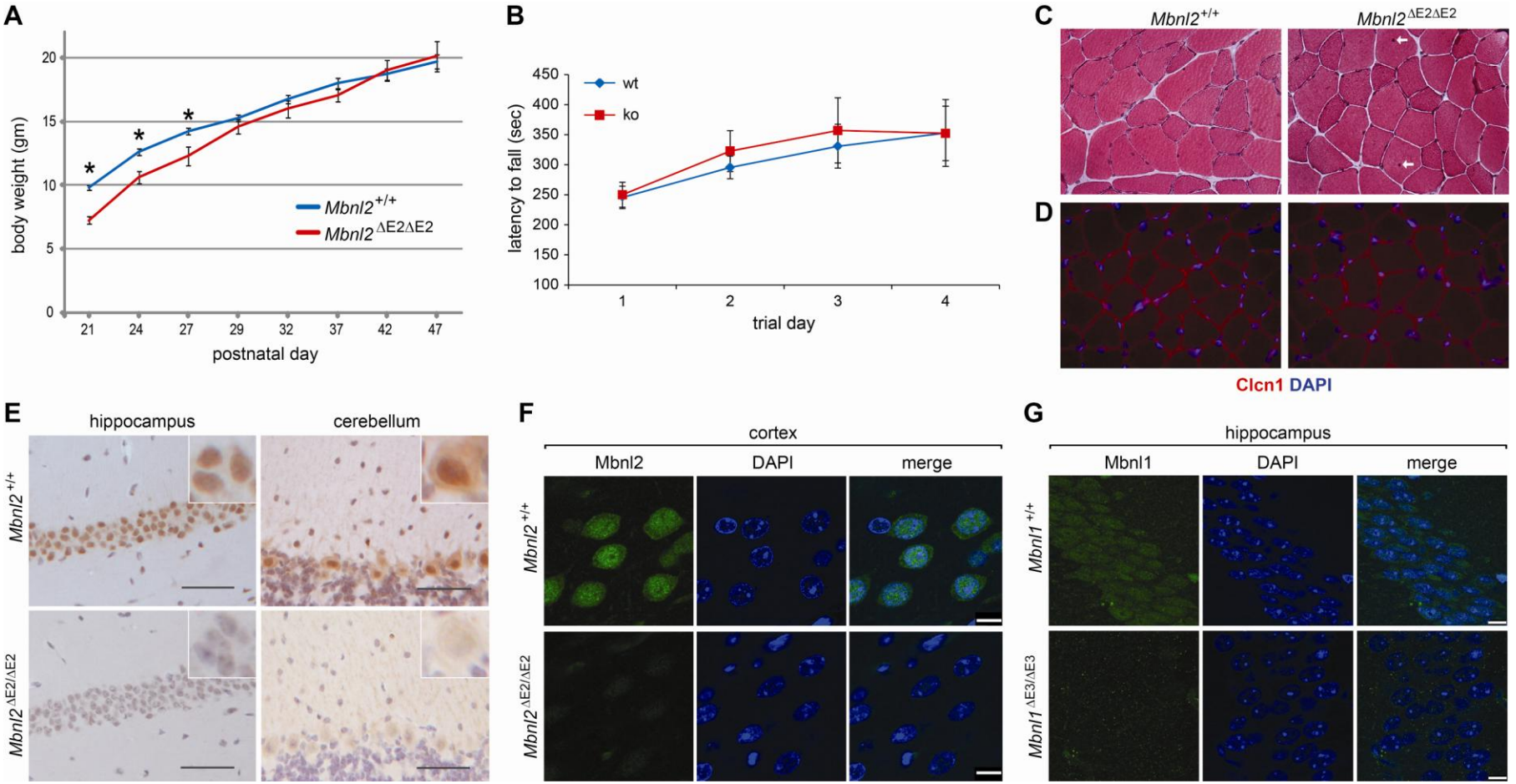

Figure S2. *Mbnl2* Knockout Growth, Absence of Muscle Pathology and *Mbnl2* Expression in the Brain

(A) *Mbnl2*$^{\Delta E2/\Delta E2}$ knockouts (red) are small at weaning compared to wild type mice (blue) but recover by P29 (**P*<0.05).

(B) Rotarod analysis of wild type (blue) versus *Mbnl2* knockouts (red). Average latency to fall is indicated over four consecutive trial days.

(C) Muscle histology (H&E staining) *Mbnl2*$^{+/+}$ wild type versus *Mbnl2*$^{\Delta E2/\Delta E2}$ knockout mice. *Mbnl2* knockout quadriceps muscle appeared normal although a 1-2 central nuclei (white arrows) were detectable in transverse sections.

(D) Frozen sections (10 μm) of quadriceps immunostained (red) using an antibody against the ClC-1 C-terminus. Myonuclei are indicated by DAPI (blue) staining.

(E) Immunohistochemistry of *Mbnl2*$^{+/+}$ versus *Mbnl2*$^{\Delta E2/\Delta E2}$ knockout hippocampus and cerebellum using anti-Mbnl2 mAb 3B4 (scale bars = 50 μm). Insets show that the primarily nuclear immunostaining is absent in both regions of the *Mbnl2* knockout brain.

(F) Immunofluorescence (anti-Mbnl2 mAb 3B4) of *Mbnl2* wild type versus knockout frontal cortex showing that Mbnl2 is distributed in both nuclear and cytoplasmic compartments in cortex neurons (scale bars = 10 μm).

(G) Immunofluorescence (anti-Mbnl1 antibody A2764) of *Mbnl1* wild type versus knockout hippocampus. Mbnl1 staining is considerably less intense (exposure time was 2X greater compared to Mbnl2, Figure 2B) and less of the signal was localized to the nucleus compared to Mbnl2 (see Figure 2B) (scale bars = 10 μm).

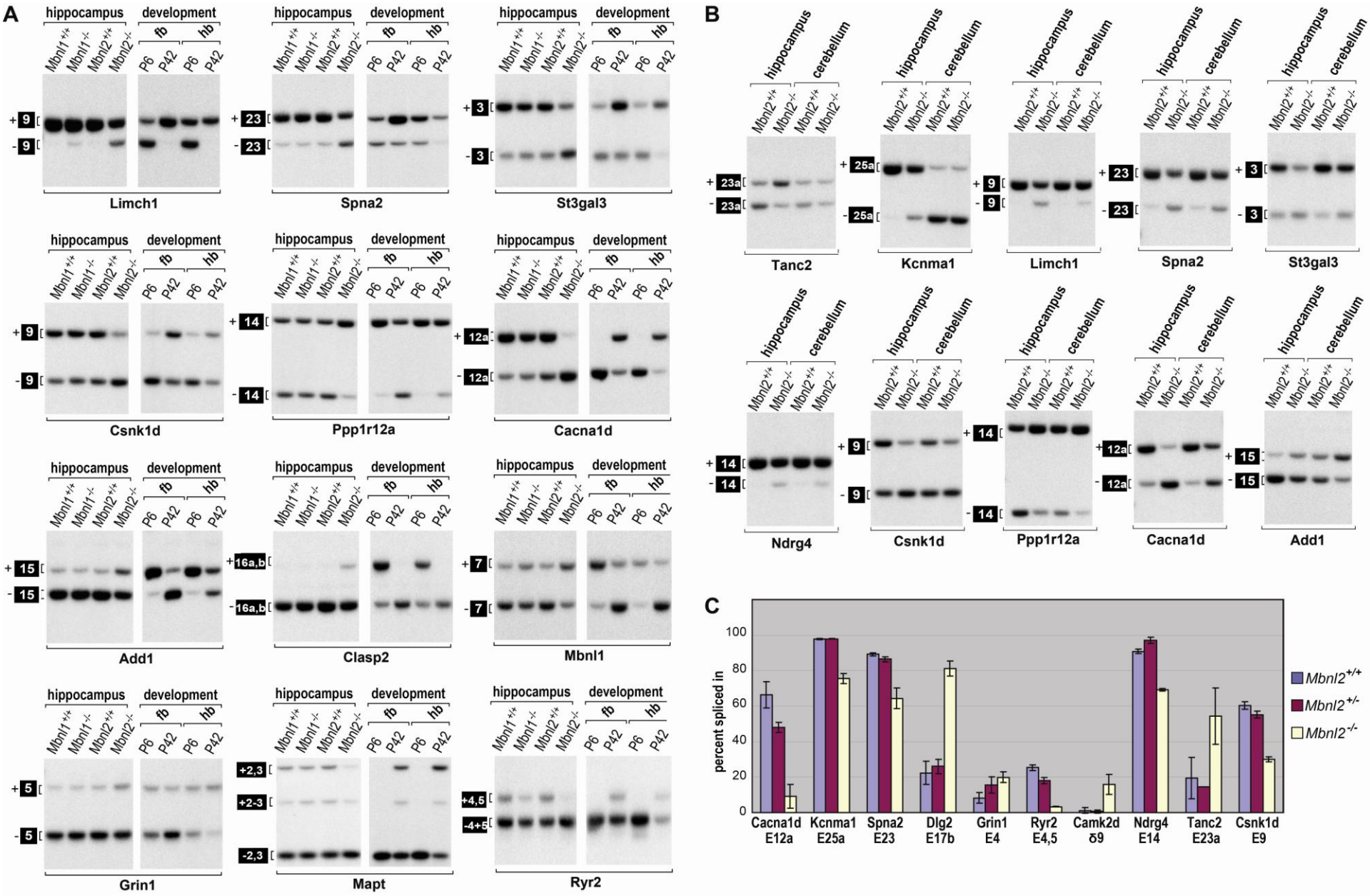

Figure S3. Development Regulation of Mbnl2 Splicing Targets in the Brain

(A) Mbnl2 targets are dysregulated in *Mbnl2*, but not *Mbnl1*, knockout mice. Splicing of indicated alternative cassettes was assayed by RT-PCR.

(B) Differential regulation of Mbnl2 splicing targets in the hippocampus and cerebellum.

(C) Genes implicated in epilepsy are mis-spliced in *Mbnl2* knockouts. Note that several targets are mis-spliced even in *Mbnl2*$^{+/\Delta E2}$ heterozygotes.

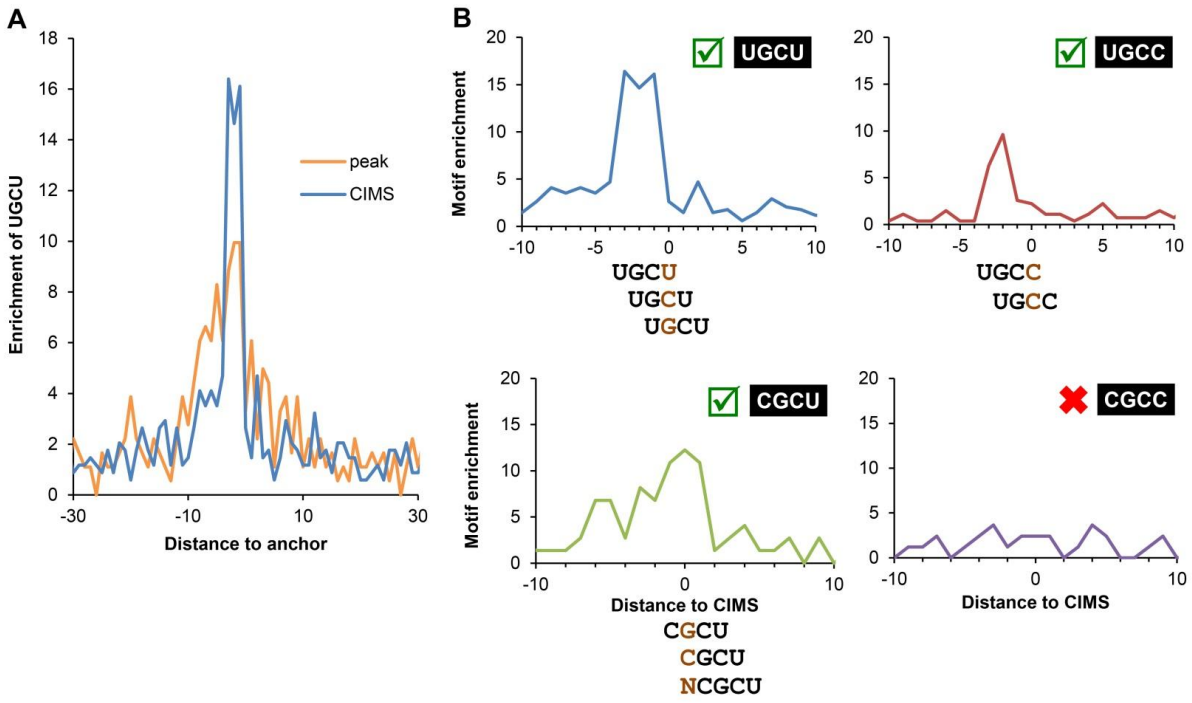

Figure S4. CIMS Analysis of Mbnl2 CLIP Data Reveals Mbnl2 Binding Specificity

(A) UGCU element is enriched around CIMS. The UGCU frequency starting at each position relative to CIMS is normalized by background motif frequency estimated from flanking sequences (−500 to −401 nt and 398–497 nt around CIMS) to obtain the normalized motif frequency. The motif enrichment around the peaks of the same CLIP tag clusters that harbor CIMS is shown as a base line for comparison.

(B) The motif enrichment around CIMS (-10 to +10 nt) is shown separately for each tetramer conforming to the YGCY consensus. Different levels of enrichment were observed for UGCU, UGCC, and CGCU, but not for CGCC. The crosslinked nucleotides can be U, C or G in the YGCY element.

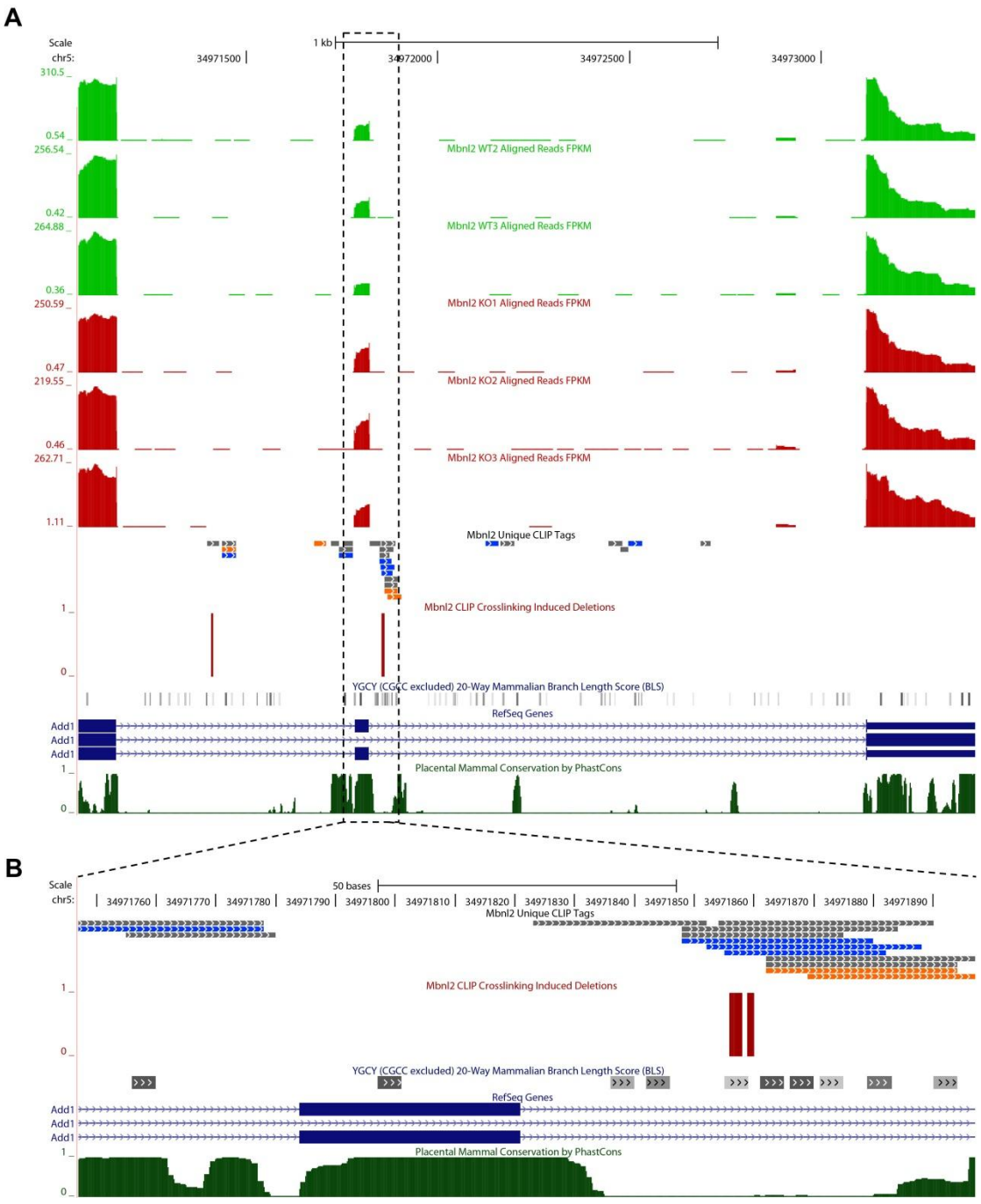

Figure S5. Mbnl2 Directly Regulates Splicing of an Add1 Cassette Exon

(A) UCSC browser view of the Add1 cassette exon. From top to bottom are the scale and genomic locus coordinates, RNA-seq read coverage quantified as FPKM (fragments per kb per million) for three wild type (green) and *Mbnl2* knockout (red) mice, respectively, Mbnl2 unique CLIP tags with each color representing a biological replicate, potential crosslinking induced nucleotide deletions (red), YGCY elements (with CGCC excluded), exon or intron structures of RefSeq transcripts, and cross-species conservation (phastCons scores) in mammals.

(B) Zoom-in view of the cassette exon and flanking intronic sequences showing unique CLIP tags, deletion sites, YGCY elements relative to the cassette. Three deletion sites close to each other, along with two overlapping with a YGCY element, were detected.

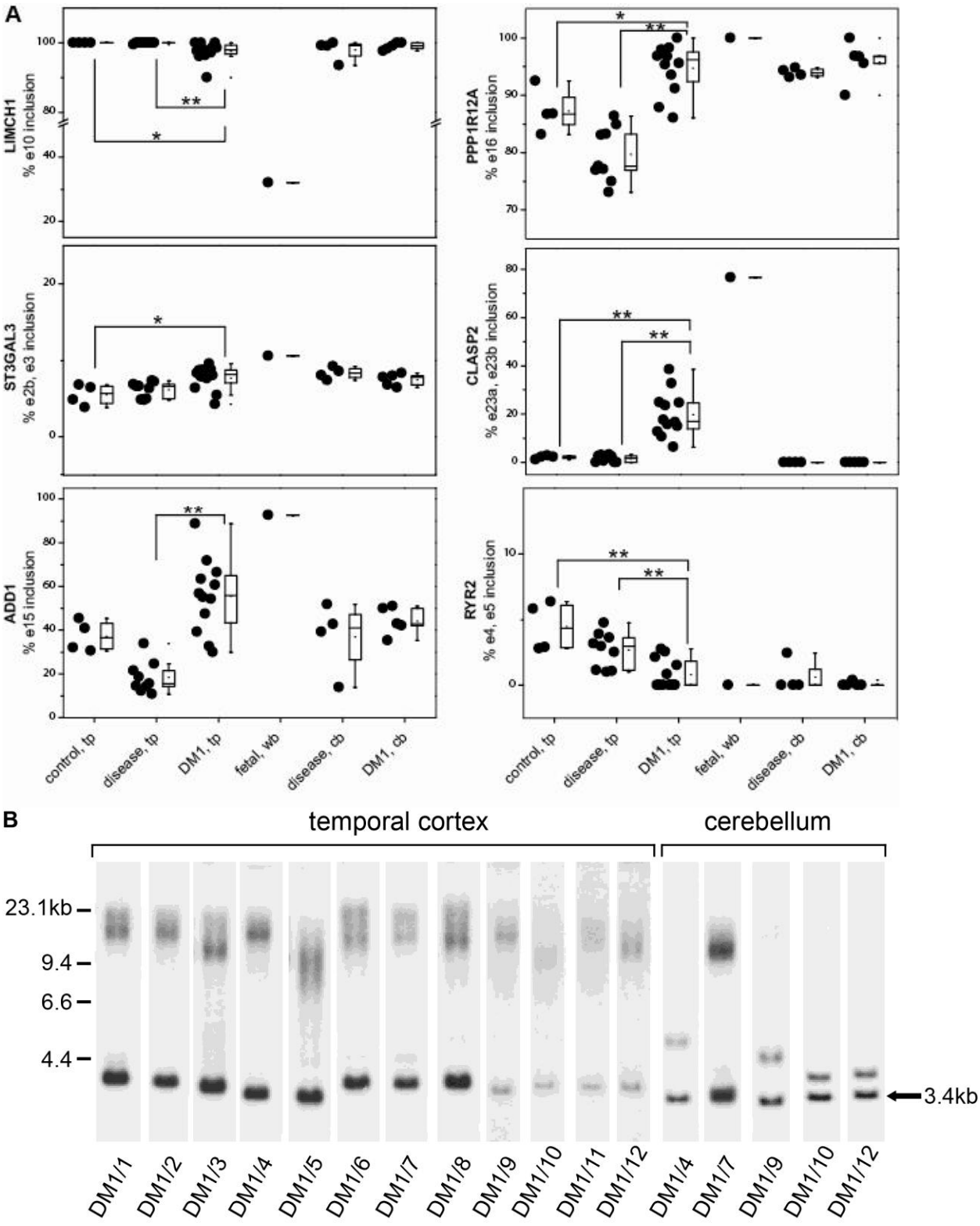

Figure S6. Mbnl2 Regulated Exons are Dysregulated in Human DM1 Brain
(A) Additional mis-spliced exons in human DM1 temporal lobe (tp) and cerebellum (cb) from control, DM1 or fetal brains. The Mann-Whitney U test was used for calculating the *P* values (**P* <0.05, ***P* <0.01).
(B) Temporal cortex (left) showed greater CTG expansions and larger CTG length heterogeneity than cerebellum (right panel). The arrow indicates the normal *DMPK* allele.

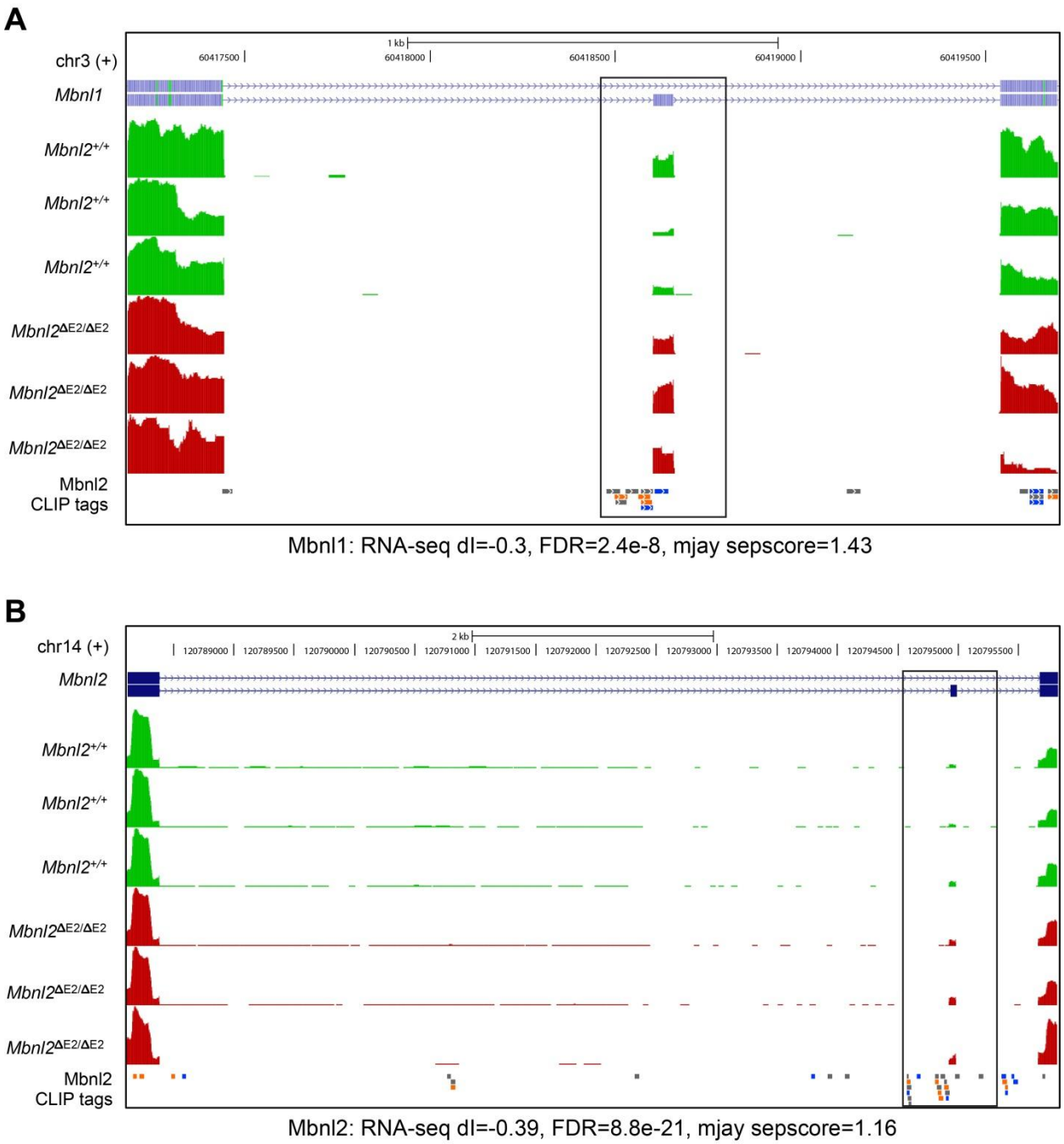

Figure S7. Cross- and auto-regulation of alternative splicing of a 54 nt exon in the *Mbnl* family
(A) Cross-regulation of *Mbnl1* exon 7 (boxed region) by Mbnl2. UCSC browser view showing that *Mbnl1* exon 7 (54 nt) splicing increases in *Mbnl2* knockouts. From top to bottom are the scale and genomic locus coordinates, RNA-seq read coverage quantified as FPKM for three wild type (green) and *Mbnl2* knockout (red) mice, respectively. Mbnl2 unique CLIP tags (each color represents a biological replicate) are also shown. Note that Mbnl2 binds directly upstream of exon 7 to repress splicing of this exon in wild type hippocampus.
(B) Autoregulation of Mbnl2 exon 6 (54 nt) by Mbnl2. Same as (A) except that Mbnl2 exon 6 (54 nt) is shown with enhanced splicing of this exon in Mbnl2 knockouts.

Table S1. Splicing microarray analysis of *Mbnl2*$^{+/+}$ versus *Mbnl2*$^{\Delta E2/\Delta E2}$ knockout hippocampus.

Splicing microarray summary of alternative cassette exons mis-spliced in *Mbnl2* knockouts.

Table S2. RNA-Seq, Gene Ontology, HITS-CLIP and CIMS summary of *Mbnl2*$^{+/+}$ versus *Mbnl2*$^{\Delta E2/\Delta E2}$ knockout hippocampus.

(A) RNA-seq summary.

(B) RNA-seq targets with 209 high-confidence Mbnl2-dependent cassette exons identified from RNA-seq (junction coverage ≥20, FDR ≤0.05, $|\Delta I| \geq 0.1$).

(C) Joint target alternative splicing including 290 combined high-confidence Mbnl2-dependent cassette exons annotated in our alternative splicing database (AS db).

(D) Mjay targets not included in AS db.

(E) Joint target genes including 271 genes with high-confidence Mbnl2-dependent splicing identified by splicing microarray or RNA-seq.

(F) Gene Ontology analysis using 271 Mbnl2 target transcripts compared to all genes expressed in hippocampus. Terms with Benjamini FDR ≤0.05 are listed.

(G) CLIP tag summary (reads, mapping and mutation statistics).

(H) List of significant Mbnl2 CLIP tag cluster peaks (P <0.01, Bonferonni correction).

(I) List of robust CIMS (FDR ≤0.001, m ≥4).

Table S3. PCR primers for *Mbnl2*$^{\Delta E2/\Delta E2}$ mouse knockout generation.

Table S4. RT-PCR primers for mouse brain Mbnl2 splicing targets.

Table S5. Summary of human autopsy samples.

Table S6. RT-PCR primers for human brain Mbnl2 splicing targets.

Table S3

PCR Primers for *Mbnl2*$^{\Delta E2/\Delta E2}$ Generation

Gene	Forward exon	Reverse exon	Product size (bp)	Forward primer	Reverse primer
Mbnl2	e2	e3	600	cagtcaagagactagaaccctgg gagc	ggcgttcctggaaacataaa
Ppia	e4	e4	142	gcggcaggtccatctacg	gccatccagccattcagtct
Mbnl1	e3	e4	308	gttagtgtcacaccaattcgggacac	gggcatcatggcattggctaac
Mbnl2$^{+/con}$	Neo, e2	i2	525 216	cgccttctatcgccttcttgacgagttcttctgag, gtagggctctcaaggagagcactgcattgagc	aatgtcaaaccagaccagaaatacaccaccatg
Mbnl2$^{+/\Delta E2}$	i1, e2	i2	188 216	gtaccaccttccttgtgatactgaaagctctgaggtc, gtagggctctcaaggagagcactgcattgagc	aatgtcaaaccagaccagaaatacaccaccatg
Mbnl2 BAC retrieval	PL253	PL253	4,900	agcctgccgtgagagagtgaagtcatcagcctccagc cacctgacttccgcggacagtggtctgctctcccagcca gggttttcccagtcacgacgttgt	ttggctcacctcccacctttacctgctttatgtcattttccgtat aaaaaggggaacgctcctccctcgttacccaacttaatc gccttgcagcacatcc
5’ LoxP targeting	PL452	PL452	1235	tcctctaagtacagacgacgcaagtcgatgtcatactgtt ttaggcctataacttcgtataatgtatgctatacgaagttat cgacctgcagcctgttga	ggtgccaatgagagtgatctccgggacaactttaaacca caaacctgcataacttcgtatagcatacattatacgaagtt atgtcgaggctgatcagcga
3’ LoxP targeting	PL451	PL451	2350	aggtttatgttgtcttttggtttggttttcactgaaacttatttat acgaagttcctattctctagaaagtataggaacttcaggt ctgaagaggagttt	ttaatttaaacacagatatggaaaaaaaaatttttgtatgc ctaagtactataacttcgtatagcatacattatacgaagtt atattatgtacctgactg
5’ probe Southern	i1	i1	325	ctctcctcctccagtttggctttg	gtgttcagacgtgagccaagactgtaac
3’ probe Southern	i2	i2	250	tcgacttcctccattctgggagaac	gcatccaggggacaattcacataga

Table S4
Mouse Brain RT-PCR Primers for Mbnl2 Target Genes

Gene	Exon	Size (bp)	Forward primer	Reverse primer
Tanc2	23a	30	gccatgattgagcatgttgactacagt (in exon 22: 133 bp)	cctcttccatcagcttgctcaaca (in exon 23b: 92 bp)
Kcnma1	25a	81	gattcacacctcctggaatggacagat (in exon 24: 114 bp)	gtgaggtacagctctgtgtcagggtcat (in exon 25b: 131 bp)
Limch1	9	36	cggaagttgccagatgtgaagaaa (in exon 8: 253 bp)	cctcctcacaccgcatgtcaaa (in exon 10: 127 bp)
Clasp2	16a,16b	27_27	gttgctgtgggaaatgccaagac (in exon 15: 102 bp)	gctccttgggatcttgcttctcttc (in exon 16c: 171 bp)
Spna2	23	60	gattggtggaaagtggaagtgaatgac (in exon 22: 149 bp)	tgatccactgctgtaactcgtttgct (in exon 24: 199 bp)
St3gal3	3	48	gcctcttcctggtcctgggattt (in exon 2: 145 bp)	caggaggaagcccagcctatcatact (in exon 4: 43 bp)
Ndrg4	14	39	cttcctgcaaggcatgggctaca (in exon 13: 52 bp)	gggcttcagcaggacacctccat (in exon 15: 1942 bp)
Csnk1d	9	63	gatacctctcgcatgtccacctcaca (in exon 8: 140 bp)	gcattgtctgcccttcacagcaaa (in exon 10: 2166 bp)
Ppp1r12a	14	171	caagcaccacatcaacaccaacagtt (in exon 13: 168 bp)	cttcgtccctaacaggagtgaggtatga (in exon 15: 91 bp)
Cacna1d	12a	60	catgcccaccagcgagactgaa (in exon 11: 88 bp)	caccaggacaatcaccagccagtaaa (in exon 13: 161 bp)
Add1	15	37	ggatgagacaagagagcagaaagagaaga (in exon 14: 99 bp)	ctgggaaggcaagtgcttctgaa (in exon 16: 1836 bp)
Mbnl1	7	54	ggctgcccaataccaggtcaac (in exon 6: 258 bp)	gggagaaatgctgtatgctgctgtaa (in exon 8: 154 bp)
Grin1	5	63	tcatcctgctggtcagcgatgac (in exon 3:177 bp)	agagccgtcacattcttggttcctg (in exon 5: 101 bp)
Camk2d	14b, 15,16	33_60_42	cagccaagagtttattgaagaaaccaga (in exon 14a: 38 bp)	ctttcacgtcttcatcctcaatggtg (in exon 17: 49 bp)
Mapt	3,4	87_87	aagaccatgctggagattacactctgc (in exon 2: 113 bp)	ggtgtctccgatgcctgcttctt (in exon 5: 66 bp)
Ryr2	4,5	21_15	cggacctgtctatctgcacctttgt (in exon 3: 105 bp)	cataccactgtaggaatggcgtagca (in exon 6: 75 bp)
Dlg2	17b	42	ccattctacaagaacaaggagcagagtga (in exon 17a: 100 bp)	gcctcgtgacaggttcataggaaaga (in exon 18: 51 bp)

Table S5

Human Autopsy Samples

Patient no/sex	Age	Disease	Sources	[d]CTG repeats in brain samples
C1/M	23	[a]normal	temporal cortex	
C2/M	22	[a]normal	temporal cortex	
C3/M	26	[a]normal	temporal cortex	
C4/F	78	[b]normal	temporal cortex	
D1/F	69	PD	temporal cortex	
D2/F	84	MSA	temporal cortex	
D3/F	80	ALS	temporal cortex	
D4/M	53	ALS	temporal cortex	
			cerebellum	
D5/F	81	ALS	temporal cortex	
			cerebellum	
D6/M	70	ALS	temporal cortex	
D7/F	75	ALS	temporal cortex	
D8/F	61	ALS	temporal cortex	
			cerebellum	
D9/M	70	ALS	temporal cortex	
			cerebellum	
DM1/M	58	DM1	temporal cortex	3000-6000
DM2/M	69	DM1	temporal cortex	3000-6000
DM3/M	50	DM1	temporal cortex	2000-6000
DM4/F	56	DM1	temporal cortex	3000-5000
			cerebellum	400
DM5/M	63	DM1	temporal cortex	1200-3000
DM6/M	59	DM1	temporal cortex	3000-9000
DM7/F	58	DM1	temporal cortex	3000-6000
			cerebellum	2200-3000
DM8/F	58	DM1	temporal cortex	2000-8000
DM9/F	58	DM1	temporal cortex	2200-6500
			cerebellum	300
DM10/F	66	DM1	temporal cortex	1900-3300
			cerebellum	150
DM11/M	64	DM1	temporal cortex	2400-5200
DM12/F	73	DM1	temporal cortex	1900-4800
			cerebellum	200
F1/F	21w	[c]fetal	whole brain	

[a]from Biochain

[b]from Ambion

[c]from Stratagene

[d]genomic DNAs from each brain region were digested with *Bgl*I and CTG repeat sizes determined by Southern blotting using a ^{32}P-labeled *DMPK* probe corresponding to sequence 5' of the CTG repeat (Accession L08835).

PD: Parkinson disease

MSA: multiple system atrophy

ALS: amyotrophic lateral sclerosis
DM1: myotonic dystrophy type1

Table S6

Human Brain RT-PCR Primers for MBNL2 RNA Targets

Gene	Exon (NCBI)	Orthologous mouse exon	Size(bp)	Forward primer	Reverse primer
TANC2	22a*	23a	30	gccatgatcgagcacgttgactacagt	cctcttccatcagcttgctcaaca
KCNMA1	27a*	25a	81	ggttcacacctccaggaatggatagat	gtgaggtacagttctgtatcagggtcat
LIMCH1	10	9	36	cggaagctgccagatgtgaagaag	cctcctcacaccgcatgtcaaa
CLASP2	23a*, 23B*	16a, 16b	27,27	gctggcatgggaaatgccaaggc	gctccgtggtatcttgcttcttttt
SPTAN1	23	23	60	gattggtggaaagtggaagtgaacgat	tgatccattgctgtagttcattcgct
ST3GAL3	2b, 3	3	48	gcctctttctggtactgggattt	caggaggaagcccaaccgatcatact
NDRG4	20	14	39	cttcctgcaaggcatgggctaca	gggcttcaacaggacacctccat
CSNK1D	9	9	64	gatacctctcgcatgtccacctcaca	gcattgtctgcccttcacagcaat
PPP1R12A	16	14	177	caagcaccacatcaacaccaacagtt	cttcatccctaacaggagtgaggtatga
CACNA1D	12	12a	60	catgcccaccagcgagactgag	caccaggacgataaccagccagtaaa
ADD1	15	15	34	ggacgaggctagagaacagaaagaaaaga	ttgggaagccgagtgcttctgaa
RYR2	4,5	4,5	21, 15	cagacctctccatctgcacctttgt	catgccactataggaatggcgcagca

* no NCBI annotation

Supplemental Experimental Procedures

Tissue Protein Analysis

Tissue lysates were obtained from *Mbnl2*$^{+/+}$ *and Mbnl2*$^{\Delta E2/\Delta E2}$ female mice (2-3 months of age) by homogenization in lysis buffer (20 mM HEPES-KOH, pH 8.0, 100 mM KCl, 0.1% Igepal CA-630 (Sigma), 0.5 mM phenylmethysulfonyl fluoride, 5 μg/ml pepstain A, 1 μg/ml chymostatin, 1 mM ε-aminocaproic acid, 1 mM *p*-aminobenzamidine, 1 μg/ml leupeptin, 2 μg/ml aprotinin) on ice, followed by sonication and centrifugation (16,100*g*, 15 min, 4°C). Mbnl2 protein was detected by immunoblotting (50 μg lysate) using mAb 3B4 (Santa Cruz Biotechnology) and HRP-conjugated anti-mouse followed by ECL (GE Healthcare). Gapdh was used as a loading control (mAb 6C5, Abcam).

Rotarod Analysis

Mice (4-5 months of age, n=5 per group) were allowed to acclimate in a behavioral facility for 1 hr and then placed on an accelerating rotarod (Accuscan Instruments) rotating at 4 RPM which was gradually increased to 40 RPM over 5 mins and then continued at 40 RPM for an additional 5 min. Latency to fall (in

seconds) from the rotating bar was recorded. Mice were rested for 10 min after each trial which was repeated three additional times per day for four consecutive days. A mouse's latency to fall for each day was recorded as the mean latency of the four consecutive trials and data are reported ± SEM.

Immunohistochemistry and X-Gal Staining

For immunohistochemistry, *Mbnl2*$^{+/+}$ *and Mbnl2*$^{\Delta 2E/\Delta E2}$ (5 months) brains were fixed overnight at 4°C in 4% paraformaldehyde (PFA) in PBS. Vibratome sections (50 µm) were processed as described (Emamian et al., 2003) and then incubated with anti-Mbnl2 mAb 3B4 in blocking solution for 72 hr followed by washing (4X, 20 min/each) in PBS and incubation with a labeled anti-mouse antibody (AlexaFluor 488, Molecular Probes) for 48 hr at 4°C. Immunohistochemistry for Clcn1 was performed as described (Kanadia et al., 2006).

To determine the *Mbnl2* spatial expression, *Mbnl2*$^{+/GT4}$ (3-5 months) brains were dissected and incubated overnight at 4°C in 0.2% PFA followed by 30% sucrose overnight at 4°C and embedded in OCT (Tissue-Tek). Transverse cryostat sections (15 µm) were dried (30 min, RT) and incubated overnight in X-

Gal staining solution (0.1M sodium phosphate, pH 7.4, 0.1% sodium deoxycholate, 2 mM MgCl, 0.2% NP-40, 1 mg/ml X-Gal, 0.1 mM $K_3Fe(CN)_6$, 0.1 mM $K_4Fe(CN)_6$) at 37°C in a humidified chamber. Stained sections were further fixed in 2% PFA (10 min, RT) and counterstained with eosin.

Mouse RNA and Splicing Analysis

RNA was isolated from *Mbnl2*$^{+/+}$, *Mbnl2*$^{\Delta E2/\Delta E2}$, *Mbnl1*$^{+/+}$ and *Mbnl1*$^{\Delta E3/\Delta E3}$ (2-3 months) dissected hippocampi and cerebella. For developmental analysis, forebrain and hindbrain were obtained from P6 and P42 wild type mice. RNAs were isolated, and cDNAs generated and PCR-amplified, as described (Kanadia et al., 2006) using specific primers (Table S4). PCR products were resolved on 10% polyacrylamide gels and band intensities quantified using a Typhoon 9200 imager and ImageQuant TL software (GE Healthcare). Unpaired Student's t test was used for the statistical analysis.

Electrophysiology

Hippocampal slices were incubated in a holding chamber (room temperature) containing standard artificial cerebrospinal fluid (ACSF): 124 mM NaCl, 2 mM KCl, 1.25 mM KH_2PO_4, 1.5 mM $MgSO_4$, 2.4 mM $CaCl_2$, 26 mM $NaHCO_3$ and 10 mM glucose. Prior to recording (30-60 min), 2-3 slices were transferred to a standard interface recording chamber (Harvard Apparatus) and the chamber was continuously perfused with standard oxygenated (95% O2, 5% CO_2) ACSF at a flow rate of 2 ml/min. The pH and temperature were maintained at 7.4 and 30 ± 0.5^{o}C, respectively. Humidified air (95% O_2, 5% CO_2) was continuously blown over the slices. Extracellular synaptic field potentials from CA3-CA1 synaptic contacts were recorded with glass micropipettes (4-6 MΩ) filled with recording medium (ACSF). Two concentric bipolar stimulating electrodes (outer pole: stainless steel, 200 µm diameter; inner pole: Platinum/Iridium, 25 µm diameter, FHC) were positioned approximately 1 mm from either side of the recording electrode localized in the middle of the stratum radiatum. A single diphasic stimulus pulse of 100 µsec was delivered from a stimulator (SD 9 Stimulator, Grass Instrument Co) to the Schaffer collateral commissural pathway, to evoke field potentials at 0.033 Hz. The signals were amplified, filtered between 1 Hz and 1 kHz, and stored for off-line analysis. Two cursors

were placed around the initial descending phase of the excitatory post synaptic potential (EPSP) waveform, and the maximum slope (mV/ms) of the EPSP was determined by a computer algorithm that found the maximum change across all sets of 20 consecutively recorded points (20 kHz sampling rate) between the two cursors. For examination of synaptic plasticity, responses were collected for at least 10 min prior to pattern stimulation to insure a stable baseline before induction of synaptic plasticity. LTP was induced by employing 4 X 1 sec trains of 100 Hz with each train 10 sec apart. Changes in transmission properties induced by patterned stimulation were calculated as the percent change from the averaged response collected during baseline. To obtain the N-methyl-D-aspartate receptor (NMDAR)-mediated component of the field EPSP, slices were incubated in ACSF containing low extracellular Mg^{2+} (0.5 mM), 6,7-dinitroquinoxaline-2,3-dione (DNQX, 30 μM), and picrotoxin (PTX, 10 μM). Input-output curves for the NMDAR response and fiber volley were constructed by measuring the maximum amplitude of the responses for increasing stimulation intensities. Analyses of variance (ANOVAs) were performed using StatView 5.0 (SAS Institute).

Splicing Microarrays and Spatial Analysis of Motif Enrichment

Hippocampal RNAs were obtained from *Mbnl2*$^{+/+}$ and *Mbnl2*$^{\Delta E2/\Delta E2}$ mice and amplified, biotinylated cDNA target was produced using the Ambion WT Expression Kit and the GeneChip Whole Transcript Terminal Labeling kit according to the manufacturer's instructions (Affymetrix). Each sample target was hybridized overnight to a Mouse GeneSplice Array (Affymetrix PN 540092). Hybridized arrays were processed using the Affymetrix Fluidics Station 450 and scanned with an Affymetrix GeneChip scanner. Data were analyzed as previously described (Sugnet et al. 2006), with some modifications. Analysis was done by calculating the sepscore which measures the relative change in ratio of alternative splicing events: sepscore = Log_2[(include/skip)mut/(include/skip)wt]. Sepscores were deemed significant if the q-value was $\leq$0.05.

Given a set of exons that showed a significant change in splicing in *Mbnl2* knockout mice, we measured enrichment of the putative *Mbnl2* recognition motif (YGCY) in the introns proximal to those exons as follows. We measured the frequency of YGCY motifs in 50-nucleotide windows that were slid along the introns upstream and downstream of this exon set, with a

sampling interval of 5 nucleotides. At each sampling interval, we computed the average number of YGCY motifs in exons preferentially included in *Mbnl2* knockdown experiments, exons preferentially skipped in the same experiments, and background exons that were scored in the same experiment but showed no change in splicing. Also in each interval, we empirically estimated the 95% confidence interval of YGCY frequency in the background set. These confidence intervals are shown as error bars in Figure 3B. Individual points outside the error bars have a YGCY frequency that is different from background at the 95% level of confidence. For peaks indicated as consecutive points outside the 95% confidence interval, we estimated the significance of each peak as follows. For each point within each peak, we applied the Mann-Whitney-Wilcoxon test to estimate a p-value that the YGCY frequency at that point is greater than background. Because the points were estimated from overlapping windows, they are not independent. Accordingly, we estimated a q-value for each peak by finding the most significant p-value for any point within the peak, and controlling for multiple hypotheses by applying Bonferroni correction with the number of points within the peak. While this correction controls for the number of points within a peak, it does not control for the number of positions at

which a peak might occur: given a peak of *M* points in a range of *N* points, there are *N – M +1* positions where the peak might occur. Thus, we applied a second Bonferroni correction to control for this number of possible positions: *N – M + 1*. This yielded the final q-values, which estimate the significance of observed peaks with appropriate controls for multiple comparisons.

Bioinformatic Analysis of Paired-end RNA Sequencing Data

Paired ends could be located in different exons and since the reads are relatively short, the transcript structure between the sequenced ends could be ambiguous when alternative splicing occurs. The missing information, leveraging on the size constraints of each cDNA fragment, was inferred using a simple Bayesian model. Specifically, the distribution of fragment size was estimated using paired-end reads located in the same exons. Large exons (≥ 400 nt) were used for this purpose to avoid bias, since the rough estimate of the average fragment size is ~130 nt. In parallel, the number of observed junction reads for each exon junction was counted. During statistical inference, for each pair of reads that spans one or more exons (up to three, which is sufficient in practice given the fragment size), we enumerated all possible isoforms (paths)

P_j between the anchored ends, and estimated the probability of each isoform to be the actual origin of the paired-end reads, using the Bayes formula:

$$\Pr(P_k \mid l) = \frac{\Pr(l \mid P_k)\Pr(P_k)}{\sum_j \Pr(l \mid P_j)\Pr(P_j)} = \frac{\Pr(l_k)\Pr(P_k)}{\sum_j \Pr(l_j)\Pr(P_j)}$$

where l_j is the length of the fragment if the reads were derived from isoform j; $\Pr(l_j)$ and $\Pr(P_j)$ are the distribution of fragment size and prior distribution of each isoform, respectively. Following analyses using inferred fragments, each was assigned a probability score. This junction inference step substantially increased the effective number of fragments supporting exon junctions, especially for cassette exons, and increased statistical power in detecting splicing changes. We then counted the weighted number of exon or exon-junction fragments uniquely supporting the inclusion or skipping isoform of each cassette exon. Biological replicates of each group were pooled together. Additional filtering criteria on junction fragment coverage were applied to reduce the multiple testing (junction_in + junction_skip≥20, junction_WT+junction_KO≥20; 2,149 cassette exons remained). A Fisher's exact test was used to evaluate the statistical significance of splicing changes using both exon and exon-junction fragments, followed by Benjamini multiple testing correction to estimate the false discovery rate (FDR). In addition,

inclusion or exclusion junction reads were used to calculate the proportional change of exon inclusion (ΔI). Differential splicing events were identified by requiring FDR<0.15, or more stringent criteria (FDR <0.05 and |ΔI| ≥0.1), which defined a subset with high confidence (Table S2).

HITS-CLIP

Hippocampi were dissected from wild type mice (11-12 weeks), snap frozen in liquid nitrogen, ground to a fine powder, and crosslinked with UV-light in a Stratalinker 1800 (Stratagene). Three biological replicates were used each consisting of four hippocampi from two mice and *Mbnl2* knockout hippocampii were used as a control. For immunoprecipitation, mAb 3B4 (5 μg) and RNase A (concentrations of 55 U/ml and 0.055 U/ml for high and low RNase, respectively) were used. cDNA libraries were generated using RNA linkers and primers described for Ago CLIP (Chi et al., 2009). All sequencing was performed using an Illumina Genome Analyzer IIx. Raw reads obtained from the Illumina pipeline include a 4 nt barcode (3 random positions and a G at the 4th position) at the 5'end, followed by the actual CLIP tag of different sizes, and possibly part of the 3' adaptor sequence. They were filtered to remove low

quality reads by requiring a minimum score of 20 in barcode positions and an average score of 20 in the following 25 nucleotides (Table S2). The filtered reads were then aligned to the mouse reference genome (mm9) by novoalign (http://www.novocraft.com), allowing iterative trimming at the 3'end (Zhang and Darnell, 2011). Potential PCR duplicates, as judged from the starting position of genomic mapping and the barcode sequences, were removed to identify unique CLIP tags. This was accomplished using a statistical model that takes into consideration sequencing errors in the barcode and the copy numbers of each barcode sequences as described previously (Darnell et al., 2011), except that in this study the sequencing errors were estimated from genomic alignment, which is more accurate than iterative inference from the barcode itself, and a confidence score of Q ≥30 was used. For CIMS analysis, we found enrichment of deletions relative to insertions (0.13 deletions vs. 0.02 insertions per tag) similar to previous findings obtained from Nova and Ago CLIP data. One or more deletions were detected in 8.9% unique CLIP tags, a rate lower than Nova but comparable to Ago (Zhang and Darnell, 2011).

Normalized Complexity RNA map

The normalized complexity map of Mbnl2 CLIP tags was generated as described previously (Licatalosi et al., 2008), with slight modifications. Specifically, the number of CLIP tags in the alternative spliced region (from the mid-point of the upstream constitutive exon to the mid-point of the downstream constitutive exon) N_i of each target exon *i* was counted. Each CLIP tag in the region was then assigned a weight 1/ N_i to normalize gene expression level, and then assigned to the closest splice site (5´ splice site of the upstream exon, the two splice sites of the cassette exon, and the 3´ splice site of the downstream exon) based on the mid-point of the CLIP tag. Four sub-regions were defined accordingly based on the distance to splice sites. For each 50-nt window *k* in each sub-region, the weighted sum of CLIP tags assigned to the sub-region t_k, as well as the number of exons with assigned CLIP tags e_k, was calculated over all activated or repressed target exons, respectively. The final normalized complexity value was defined to be $s_k=100\times t_k \times e_k / E^2$, where *E* is the number of activated or repressed exons.

Gene Ontology Analysis

Gene Ontology analysis was performed using DAVID (Huang da et al., 2009). For background genes, we used the top half of the genes according to expression values estimated from RNA-seq data.

Genomic Blot Analysis

Genomic DNAs from each region of autopsied DM1 brains were digested with *Bgl*I. The CTG trinucleotide repeat expansion sizes were determined by Southern blot analysis using a ^{32}P-labeled *DMPK* probe (Mitsubishi Chemical Medience Corporation).

Supplemental References

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

Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W., *et al.* (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell *146*, 247-261.

Emamian, E.S., Kaytor, M.D., Duvick, L.A., Zu, T., Tousey, S.K., Zoghbi, H.Y., Clark, H.B., and Orr, H.T. (2003). Serine 776 of ataxin-1 is critical for polyglutamine-induced disease in SCA1 transgenic mice. Neuron *38*, 375-387.

Huang da, W, Sherman, B.T, Lempicki, R.A. (2009) Systematic and integrative analysis of large gene sets using DAVID bioinformatics resources. Nat Proc *4*, 44-57.

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.

Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X.*, et al.* (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature *456*, 464-469.

Sugnet, C.W., Srinivasan, K., Clark, T.A., O'Brien, G., Cline, M.S., Wang, H., Williams, A., Kulp, D., Blume, J.E., Haussler, D.*, et al.* (2006). Unusual intron conservation near tissue-regulated exons found by splicing microarrays. PLoS Comput Biol *2*, e4.

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