ORIGINAL ARTICLE

Extensive cryptic splicing upon loss of RBM17 and TDP43 in neurodegeneration models

Qiumin Tan[1,2,†], Hari Krishna Yalamanchili[1,2,†], Jeehye Park[1,2,†,‡],
Antonia De Maio[2,3], Hsiang-Chih Lu[2,3], Ying-Wooi Wan[1,2],
Joshua J. White[2,4,5], Vitaliy V. Bondar[1,2], Layal S. Sayegh[1,2], Xiuyun Liu[1,2],
Yan Gao[1,2], Roy V. Sillitoe[2,3,4,5], Harry T. Orr[6,7], Zhandong Liu[2,8,*] and
Huda Y. Zoghbi[1,2,3,4,5,9,*]

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA,
[2]Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Houston, Texas 77030, USA,
[3]Program in Developmental Biology, [4]Department of Neuroscience, [5]Department of Pathology and
Immunology, Baylor College of Medicine, Houston, Texas 77030, USA, [6]Institute for Translational
Neuroscience, [7]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis,
Minnesota 55455, USA, [8]Department of Pediatrics and [9]Howard Hughes Medical Institute, Baylor College of
Medicine, Houston, Texas 77030, USA

*To whom correspondence should be addressed at: Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Houston, TX 77030,
USA. Tel: +1 7137986558; Fax: +1 7137988728; Email: hzoghbi@bcm.edu (H. Y. Z.) or Tel: +1 8328248878; Fax: +1 8328251251; Email: zhandonl@bcm.edu (Z. L.)

Abstract

Splicing regulation is an important step of post-transcriptional gene regulation. It is a highly dynamic process orchestrated by RNA-binding proteins (RBPs). RBP dysfunction and global splicing dysregulation have been implicated in many human diseases, but the *in vivo* functions of most RBPs and the splicing outcome upon their loss remain largely unexplored. Here we report that constitutive deletion of *Rbm17*, which encodes an RBP with a putative role in splicing, causes early embryonic lethality in mice and that its loss in Purkinje neurons leads to rapid degeneration. Transcriptome profiling of *Rbm17*-deficient and control neurons and subsequent splicing analyses using CrypSplice, a new computational method that we developed, revealed that more than half of RBM17-dependent splicing changes are cryptic. Importantly, RBM17 represses cryptic splicing of genes that likely contribute to motor coordination and cell survival. This finding prompted us to re-analyze published datasets from a recent report on TDP-43, an RBP implicated in amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD), as it was demonstrated that TDP-43 represses cryptic exon splicing to promote cell survival. We uncovered a large number of TDP-43-dependent splicing defects that were not previously discovered, revealing that TDP-43 extensively regulates cryptic splicing. Moreover, we found a significant overlap in genes that undergo both RBM17- and TDP-43-dependent cryptic splicing repression, many of which are associated with survival. We propose that repression of cryptic splicing by RBPs is critical for neuronal health and survival. CrypSplice is available at www.liuzlab.org/CrypSplice.

Introduction

To produce functional gene products in eukaryotes, nascent RNA must undergo multiple steps of processing, including pre-mRNA splicing. A large ribonucleoprotein complex called the spliceosome assembles at the intron–exon junctions and carries out splicing reactions to remove introns and join exons, while many regulatory proteins, such as RNA-binding proteins (RBPs), modulate the splicing outcome for particular mRNAs (1,2). The splicing process is tightly regulated and is crucial for proper gene expression. It is thus not surprising that dysfunctional spliceosome or other RBPs have been implicated in human diseases ranging from cancer to neurodegeneration. For example, *SF3B1* and *U2AF35*, both encoding core spliceosome components, are frequently mutated in chronic lymphocytic leukemia and myelodysplasia (3,4). Mutations in *TDP-43*, *FUS*, *ATXN2* and *MATR3*, all of which encode RBPs, are implicated in familial amyotrophic lateral sclerosis (ALS); *TDP-43* has also been linked to frontotemporal dementia (FTD) (5–9).

The relevance of spliceosomes and regulatory RBPs to human disease has spurred tremendous interest in studying their pathogenic mechanisms. Accordingly, many studies focus on uncovering genome-wide splicing changes caused by dysfunctional RNA-processing proteins that may affect splicing. One challenge in doing so is that many of the established computational tools used for splicing analysis rely on isoform definition or exon annotations (10). However, the eukaryotic genome has a large number of cryptic splice sites, which are suboptimal splice sites that are rarely used under normal conditions. These dormant sites can be activated when the nearby strong splice site is mutated or when there are defects in RNA-processing proteins (11–14). For example, when TDP-43 is removed from mouse embryonic stem cells, the splicing of many cryptic exons was activated. This phenomenon was also observed in cells from human patients with ALS and FTD (13). Splicing analysis based on known exon annotation will miss such cryptic changes. In this case, had the authors not performed an exhaustive manual search, the role of TDP-43 in cryptic exon repression would not have been revealed. Thus there is a biological and computational need to systematically interrogate cryptic splicing events when studying RBP functions and disease pathogenesis.

One of the RBPs that has been implicated in cancer and a neurodegenerative disease, but has not been studied in great detail, is RNA-binding motif protein 17 (RBM17) (15,16). Originally identified as a component of the spliceosome complex, it interacts with splicing factors U2AF2, SF1 and SF3B1 (17–19); our lab has also discovered that RBM17 binds Ataxin-1 in a polyglutamine- and phosphorylation-dependent manner (15). Limited studies on the splicing functions of RBM17 so far have suggested a role in alternative splicing. RBM17 promotes the usage of the upstream cryptic 3' splice site AG of the alternative exon 3 in *Drosophila* sex-lethal (*Sxl*). Loss of RBM17 leads to exon 3 inclusion, compromising the autoregulation loop of *Sxl* (20). RBM17 has also been shown to regulate the alternative splicing of *FAS* as its overexpression promotes exon 6 skipping, whereas its depletion favors exon 6 inclusion (18,21). To date, whether and how RBM17 regulates genome-wide splicing remains elusive.

Here, we report that constitutive loss of *Rbm17* in mice caused embryonic lethality, and that hindbrain-specific deletion led to cerebellar and midbrain abnormalities and early death. To circumvent the premature death and to investigate the potential *in vivo* effects of RBM17 on splicing regulation, we generated a conditional mouse mutant that lacks *Rbm17* in cerebellar Purkinje neurons and studied the transcriptome-wide expression and splicing patterns. In contrast to its proposed function in
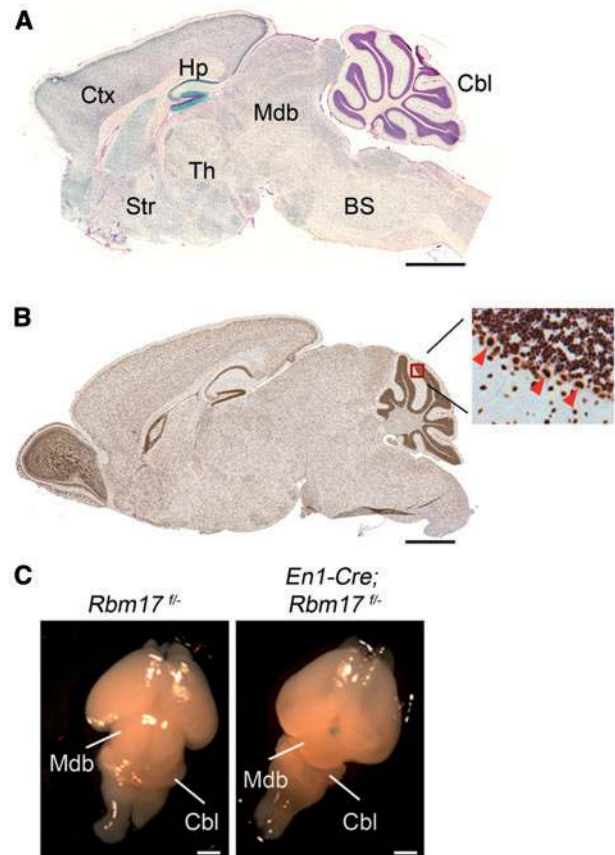

Figure 1. RBM17 is widely expressed and important for development and survival. (**A**) *Rbm17* expression in mouse brain of *Rbm17* [+/-] revealed by X-gal staining (blue). BS, brain stem; Cbl, cerebellum; Ctx, cortex; Hp, hippocampus; Mdb, mid-brain; Str, striatum; Th, thalamus. Scale bar, 2 mm. (**B**) RBM17 immunoreactivity is found throughout the brain. Scale bar, 2 mm. A region in the cerebellum is shown under higher magnification in the inset. Arrowheads indicate the nuclear staining in Purkinje neurons. (**C**) *En1-Cre; Rbm17* [f/-] mice had smaller midbrains (Mdb) and cerebella (Cbl) compared to control littermates at P0. Scale bar, 2 mm.

alternative splicing, we found that RBM17 regulates isoform splicing in a limited number of genes. This raised the possibility that RBM17 might function beyond alternative splicing and affect splicing events that are unannotated (cryptic). We developed an algorithm called CrypSplice, a novel cryptic splice site detection method and discovered that RBM17 plays a role in repressing cryptic splicing junctions. Given the similarities between RBM17 and TDP-43 in cryptic splice site repression, we re-visit the published TDP-43 RNA-seq data (13). CrypSplice identified substantially more cryptic splicing changes than reported, and revealed a common set of genes with cryptic splicing de-repression between *Rbm17* and *Tdp-43* null cells.

Results

Loss of *Rbm17* leads to abnormal development

To assess the functions of RBM17 in mice, we generated a mouse model harboring an *Rbm17* null allele in the C57BL/6 background (Supplementary Material, Fig. S1). Utilizing the *β-Gal* gene that was inserted into the endogenous *Rbm17* locus, we performed LacZ staining and found that *Rbm17* was widely expressed in the brain, including the cortex, hippocampus, midbrain, cerebellum and brain stem (Fig. 1A). Analyses of RBM17
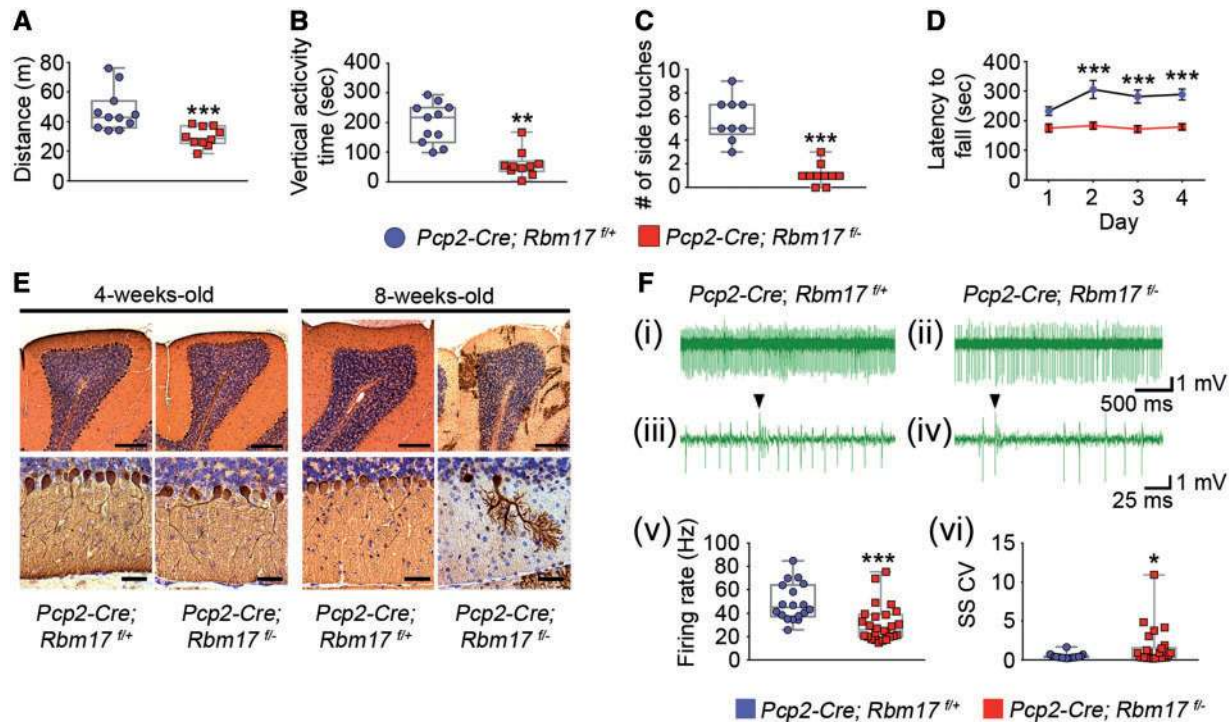
Figure 2. Loss of *Rbm17* in Purkinje cells leads to ataxia and neurodegeneration. (**A** and **B**) *Pcp2-Cre; Rbm17 $^{f/-}$* knockout mice showed reduced activity in the open field assay. (**C**) *Pcp2-Cre; Rbm17 $^{f/-}$* knockout mice had impaired motor function on the dowel test. (**D**) *Pcp2-Cre; Rbm17 $^{f/-}$* knockout mice performed worse on the rotarod test. Values were plotted with the box and whisker plots showing all data points with n = 9–11 per genotype at 8–9 weeks of age. (***$P < 0.001$; **$P < 0.01$) (**E**) Histopathological defects revealing Purkinje cell degeneration. Calbindin staining showing Purkinje cells in the *Pcp2-Cre; Rbm17 $^{f/-}$* mice when animals were at 4 and 8 weeks of age. Upper panel scale bars, 500 μm. Lower panel scale bars, 100 μm. (**F**) Electrophysiological defects in Purkinje cells in *Pcp2-Cre; Rbm17 $^{f/-}$* mice. Purkinje cells were identified by the presence of a unique action potential waveform called the complex spike (black arrowheads). Representative in *vivo* spike recordings of Purkinje cells from *Pcp2-Cre; Rbm17 $^{f/+}$* (**i** and **iii**) and *Pcp2-Cre; Rbm17 $^{f/-}$* (**ii** and **iv**) mice are shown. Quantifications of firing rate and coefficient of variation for Purkinje cell simple spikes (SS CV) are shown in **v** and **vi**. (Animal age: 4-weeks old, n = 18–26; ***$P < 0.001$; *$P < 0.05$).

protein distribution revealed a similar anatomic pattern to the LacZ staining studies with predominant staining found in the nucleus (Fig. 1B, Supplementary Material, Fig. S3A, compared with the conditional knockout), consistent with its putative function as a splicing factor. *Rbm17 $^{+/-}$* mice were healthy and viable with no obvious phenotypes. RBM17 protein levels were reduced by about 40% in the heterozygous mice (Supplementary Material, Fig. S2). We could not obtain viable *Rbm17 $^{-/-}$* offspring upon inter-crossing between heterozygous animals, and timed mating experiments showed that *Rbm17 $^{-/-}$* embryos died before embryonic day 8.5 (E8.5, Supplementary Material, Table S1A). We then resorted to the generation of tissue- and cell-specific knockouts of *Rbm17* to dissect its functions *in vivo*. When *Rbm17* was deleted from the midbrain and cerebellar primordia using *Engrailed-1 (En1)-Cre* (22), no conditional knockout pups survived beyond postnatal day 4 (P4, Supplementary Material, Table S1B). At P0, the midbrains and cerebella from *En1-Cre; Rbm17 $^{f/-}$* mice were considerably smaller than those from the controls (Fig. 1C), indicating that RBM17 is critical for the proper development of these regions.

Purkinje cell-specific *Rbm17* knockout mice develop ataxia and neurodegeneration

Given that whole-body and tissue-specific knockout of *Rbm17* resulted in early lethality (Supplementary Material, Table S1), and because RBM17 is robustly expressed in cerebellar Purkinje neurons (Fig. 1B), we next focused on this single population and

generated conditional knockout mice that lack *Rbm17* in these cells (*Pcp2-Cre; Rbm17 $^{f/-}$*); the *Pcp2-Cre* is expressed postnatally in Purkinje cells starting from day 6 (23,24). Immunohistochemistry confirmed the loss of RBM17 from Purkinje neurons in the conditional knockout mice (Supplementary Material, Fig. S3A). These mice were born at the expected Mendelian ratio and were viable, but they developed noticeable ataxia at eight weeks of age. They showed reduced open-field activity and worse performance on the dowel and rotating rod tests compared with controls (Fig. 2A–D), which suggests impaired motor coordination.

Histology and immunostaining with a Calbindin antibody showed rapid and progressive reduction of Purkinje cells in the conditional knockout mice (Fig. 2E, Supplementary Material, Fig. S3B). Of note, Purkinje cell loss was not evident at four weeks of age, but by 8 weeks of age only 10% of the expected Purkinje cells remained. Prior to this period of cell death, *in vivo* electrophysiological recordings showed reduced and irregular firing of mutant Purkinje cells in 4-week old mice (Fig. 2F).

Loss of *Rbm17* causes upregulation of apoptotic genes

Next we investigated the molecular mechanism by which loss of RBM17 leads to rapid cell death. Because RBM17 regulates the splicing of *Sxl* in *Drosophila* and *FAS* in cultured mammalian cells (18,20,21,25), we sought to assess global gene expression and mRNA splicing changes in Purkinje cells upon deletion of *Rbm17*. To this end, we used the TRAP (translating ribosome
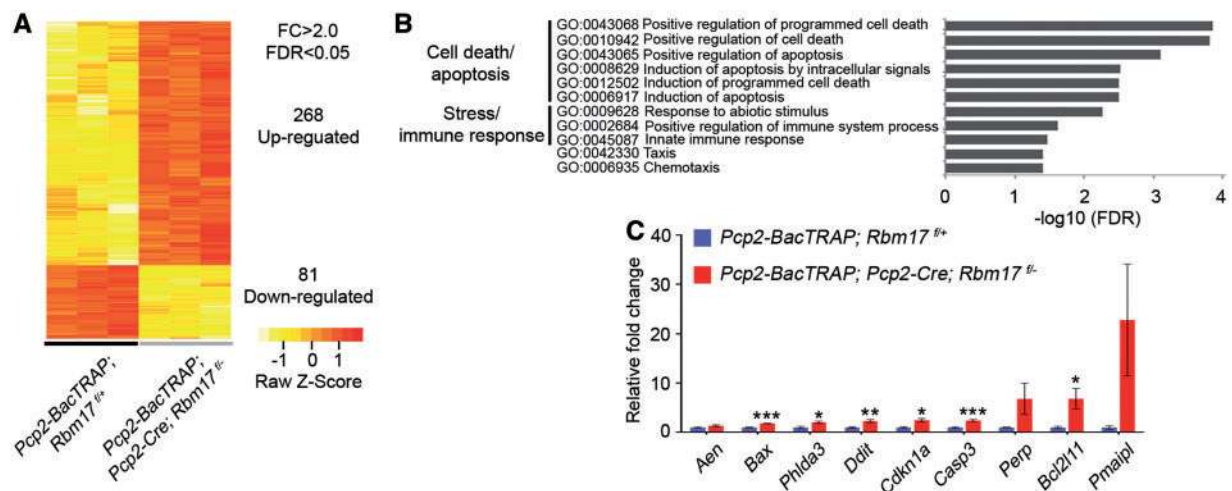
Figure 3. Gene expression changes in *Rbm17*-deficient Purkinje cells revealed by BacTRAP profiling. (**A**) Heatmap showing the number of differentially expressed genes (DEGs) using a cut-off of fold change > 2.0 and FDR < 0.05. (**B**) GO analyses of functional categories enriched with DEGs. Cell death and apoptotic pathways are the predominant pathways. (**C**) Quantitative qPCR validating the upregulated apoptotic genes. n = 3–6 animals per genotype. (***P < 0.001; **P < 0.01; *P < 0.05).

affinity purification) approach, which allows the isolation and subsequent deep sequencing of cell type-specific mRNAs that are bound to the ribosomes (26,27).

We bred mice from the *Pcp2-BacTRAP* line with the *Pcp2-Cre; Rbm17 f/−* mice to generate Purkinje cells that express the GFP-tagged ribosome subunits (Supplementary Material, Fig. S4A). TRAP was performed when animals were four weeks old, after the cerebellum is mature but before Purkinje cell loss. The enrichment of Purkinje cell-specific mRNAs and the reduction of *Rbm17* were confirmed prior to beginning RNA-seq (Supplementary Material, Fig. S4B and C). Bioinformatic analysis of the RNA-seq datasets revealed a total of 349 genes that were differentially expressed between the control and mutant Purkinje cells (fold change > 2.0, FDR < 0.05, Fig. 3A, Supplementary Material, Table S2). Gene ontology (GO) analysis showed that apoptotic genes were overrepresented among the upregulated genes, whereas ion channel genes were enriched among the most downregulated genes, as validated by qRT-PCR (Fig. 3B and C, Supplementary Material, Fig. S4D). These findings might explain the death of Purkinje cells in the absence of *Rbm17*.

CrypSplice algorithm finds previously unknown splicing changes and reveals RBM17 represses splicing

To investigate the proposed role of RBM17 in alternative splicing of isoforms, we evaluated the splicing patterns in our RNA-seq data using Cufflinks (28) and found 134 differentially expressed isoforms (Supplementary Material, Table S3). Though this finding indicates that RBM17 regulates the alternative splicing of these genes, the magnitude of the effect does not seem impressive. We then used DEXSeq (29) to search for differentially expressed annotated exons and found 708 genes where at least one of the exons was significantly different between the two genotypes (Supplementary Material, Table S3). Although this seems to suggest that RBM17 influences exon selection, it is worth noting that differentially expressed exons might not conform to differential splicing. Changes in overall gene expression can also lead to differential exon expression.

Both Cufflinks and DEXSeq are limited, however, in that they interrogate the data at the resolution of either the transcript or the annotated exon and are likely to miss functionally important junctions that have not been previously annotated. Since we know very little about the physiological functions of RBM17, we wanted to further explore the unannotated part of the genome and developed a new method, CrypSplice, to explicitly search for 'cryptic' splicing changes that would have been missed by established methods. Here, we define *cryptic* splicing junctions as the ones that have not been reported in public databases and were not found in either RNA-seq datasets from *Pcp2-BacTRAP; Rbm17 f/+* or *Pcp2-BacTRAP; Pcp2-Cre; Rbm17 f/−* neurons using standard approaches. CrypSplice takes junction counts from any popular read-mapping algorithms, such as TopHat v2.0.9 (30) in this study. It then filters out all known reported junctions and focuses on potential cryptic junctions. To minimize alignment noise, junctions with read coverage less than a user-specified threshold x (in our study 10) are ignored. Every junction in a sample is then quantified as the ratio of junction to 5′ splice site coverage and subjected to a beta binomial test (31) and corrected for multiple testing. Significant junctions with an adjusted P value < 0.01 are reported as cryptic splicing junctions. Junctions spanning more than one gene are not considered for reporting genes. A detailed description of the CrypSplice and the underlying beta binomial model is given in the materials and methods section and the workflow of CrypSplice is illustrated in Figure 4.

CrypSplice analyses revealed that loss of *Rbm17* produced 2177 cryptic splicing junctions in 1475 genes (Supplementary Material, Table S4), substantially more than those found by the established methods Cufflinks and DEXseq (134 and 708 genes, respectively). In mutant neurons, >90% of the RBM17-dependent cryptic splicing events are gain-of-cryptic junctions, suggesting RBM17 normally represses the splicing of these junctions (Supplementary Material, Table S4, Supplementary Material, Fig. S5). Of note, not all the transcripts from a given gene underwent splicing changes. To estimate the relative abundance of the abnormal transcript variants, we calculated the ratio of the junctional sequencing reads in the variant region to the total reads coming from the upstream 5′ exon in the *Pcp2-BacTRAP; Pcp2-Cre; Rbm17 f/−* cells (Supplementary Material, Table S4 and S5). We found that 88% of the aberrantly spliced transcripts contributed to <20% of the expressed transcripts (Supplementary Material, Fig. S6), which suggests that the vast majority of the cryptic splicing events would not affect the overall gene expression. This is consistent with our findings that
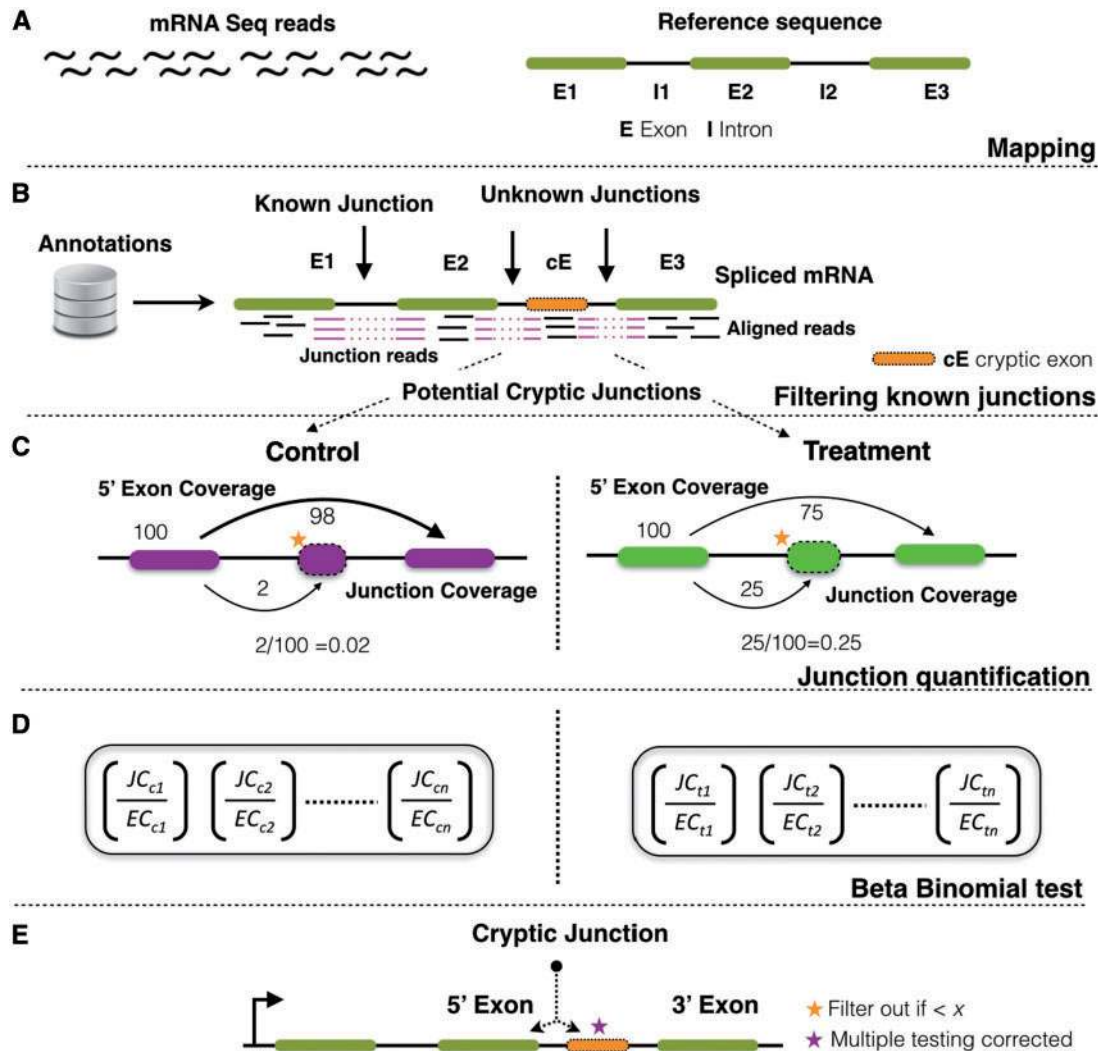
Figure 4. Workflow of CrypSplice. **(A)** Read alignment and junction quantification. **(B)** Filtering out known junctions and collapsing overlapping junctions. **(C)** Computing junction scores. **(D)** Performing beta-binomial test. **(E)** Multiple testing corrections.

only 3.45% (51 genes out of 1475 genes) with cryptic splicing changes also displayed differential expression identified by DEseq. We validated 80% (22 out of 29) of the randomly selected splicing changes using RT-PCR with three extra independent samples per genotype. We observed a wide variety of cryptic splicing junction gains, including premature 3′ UTR (Fig. 5A), cryptic exon inclusion (Fig. 5B), and exon extension (Fig. 5C, Supplementary Material, Table S5). Most of the verified splicing changes (16/22) were inclusion of cryptic exons, the majority of which (15/16) were predicted to introduce a premature termination codon (PTC).

It was previously reported that RBM17 binds to the upstream cryptic 3′ splice site AG in the presence of an intact downstream AG (20). We next examined whether there exist consensus motifs near RBM17-dependent cryptic splice sites. We took relatively strong cryptic junctions, which were incorporated into at least 10% of total transcripts, and scanned for consensus motifs in the proximity of their 5′ (upstream -100 bp to downstream 400 bp) and 3′ (upstream -400 bp to downstream 100 bp) splice sites. Motif analyses predicted more than 10 consensus motifs proximal to each site (Supplementary Material, Table S6). Some of the identified motifs were established motifs for splicing reaction. For example, the (C/A)GGUA motif is a 5′ splice donor site, and CCT(G/U)(U/C)CUC could be the pyrimidine tract. Interestingly, the top motifs surrounding both the 5′ and 3′ splice sites are rich in A, and there are three motifs with consensus AG near the 3′ splice sites. These motifs might provide the cis-acting RNA sequence context for RBM17 binding.

To gain further insight into how splicing dysregulation might contribute to the observed phenotypes in the *Rbm17* knockout mice, we searched for phenotypes associated with genes displaying splicing defects, both annotated and cryptic, and found a significant enrichment in genes associated with abnormal synaptic transmission, abnormal motor coordination and premature lethality (Supplementary Material, Fig. S7), consistent with the phenotypes that we observed in mutant mice (Fig. 1 and 2). Intriguingly, the majority of the genes associated with these phenotypes gained cryptic junctions in mutant neurons (Fig. 5D, Supplementary Material, Table S7). These findings suggest that splicing defects, especially de-repression of cryptic splicing junctions leading to the inclusion of intronic elements in mature transcripts, contribute to the observed mutant phenotypes.
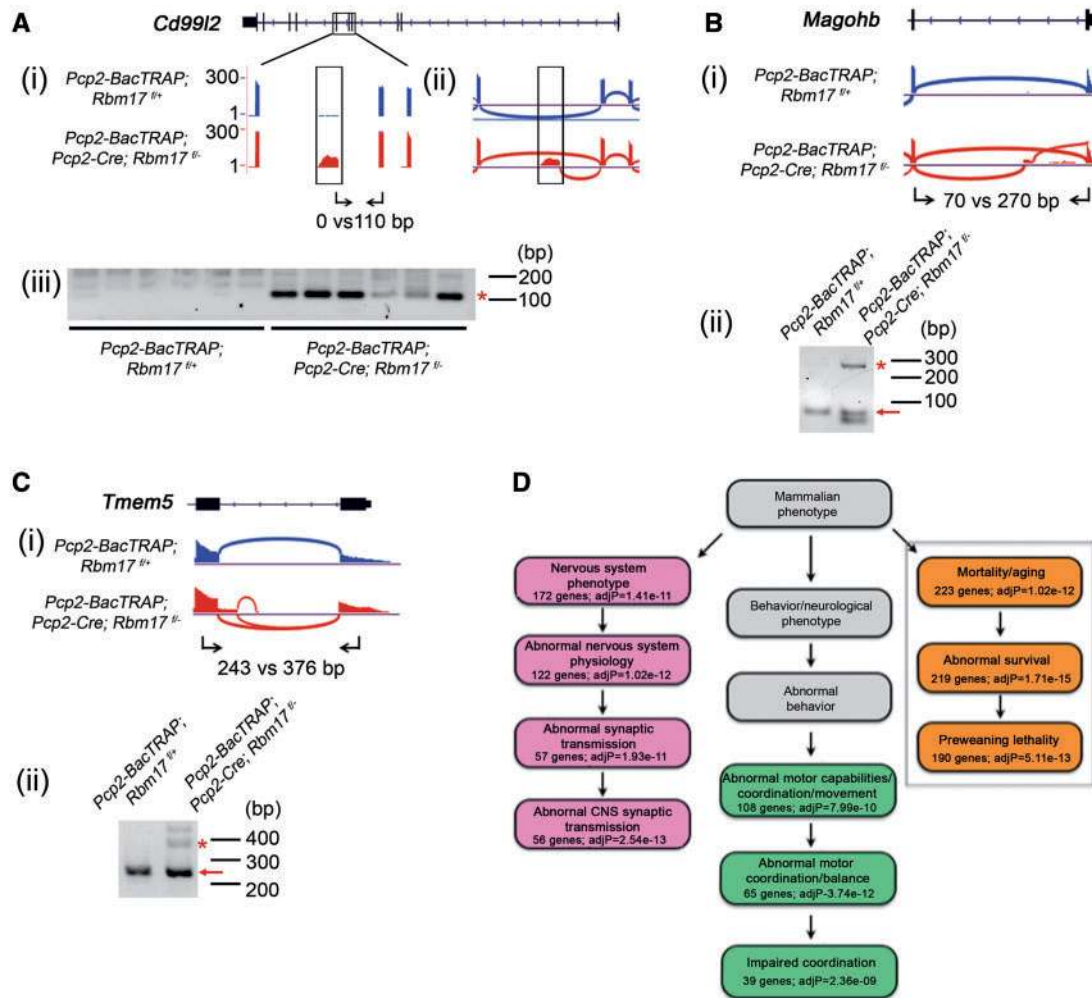
Figure 5. Cryptic splicing changes in *Rbm17*-deficient Purkinje cells revealed by BacTRAP profiling and CrypSplice analyses. (**A**) PCR validation of a premature 3′ UTR event in *Cd99l2*. The genome browser view is shown in (**i**) and the respective Sashimi plots depicting the connection between reads are shown in (**ii**). PCR gel image is shown in (**iii**) with the expected product size in the knockouts indicated by an asterisk. (**B**) PCR validation of a cryptic exon inclusion event in *Magohb*. (**C**) PCR validation of an exon extension event in *Tmem5*. Sashimi plots depicting the connection between reads are shown in (**i**). PCR gel images are shown in (**ii**) with the expected products in the controls indicated with red arrows and those in the knockouts indicated by asterisks. (**D**) Enrichment analyses of phenotypes significantly associated with genes displaying RBM17-dependent cryptic splicing repression. Phenotypes with significant enrichment are shown in colored boxes. The number of genes associated with each significantly enriched phenotype and their respective adjusted *P* values are also indicated.

Genes associated with premature death display RBM17- and TDP-43-dependent cryptic splicing repression

Our results revealed that RBM17 represses many cryptic splicing junctions, a function similar to what has been reported for TDP-43 (13). This prompted us to re-visit the published datasets on TDP-43. Mutations in *TDP-43* are linked to familial ALS (6,32), and mouse knockouts for *Tdp-43* are embryonic-lethal (33,34). A recent report showed that TDP-43 represses the splicing of cryptic exons; this activity is crucial for cell survival, as normalization of cryptic splicing activities using a splicing repressor domain fused to the N-terminus of TDP-43 rescued cell death mediated by *Tdp-43* depletion (13). Fewer than 50 targets were identified in that study, however, because the search for TDP-43-dependent splicing events was performed manually. We therefore first analyzed the published RNA-seq data from human HeLa cell lines treated with *TDP-43* siRNA using CrypSplice. We identified 1274 genes with cryptic junction gains when *TDP-43* was reduced including 33 of 34 of the previously reported genes (Supplementary

Material, Table S8). We further selected 6 previously reported and 5 newly identified cryptic splicing events and validated 5 and 3 of them, respectively, by RT-PCR (Fig. 6A and B; Supplementary Material, Fig. S8; Supplementary Material, Table S9), supporting the validity of CrypSplice. In the previous report, no overlap in TDP-43-dependent cryptic splicing events was observed between human and mouse cells deficient in *TDP-43*, likely because only a few genes (34 in human cells and 48 in mouse cells) were examined. To make a comprehensive comparison, we re-examined the RNA-seq data from Ling *et al.* (13) on Cre^ER-induced *Tdp-43* knockout mouse embryonic stem cells using CrypSplice and found 1550 genes with TDP-43-dependent cryptic splicing repression, including 42 out of the 48 from the previous report (Supplementary Material, Table S10). These results show a substantial expansion in the gene lists of TDP-43-dependent cryptic splicing and demonstrate the detection power of CrypSplice. When we compared these new lists of genes with TDP-43-dependent cryptic splicing changes in both human and mouse cells, we found an intersection of 173 genes, which are
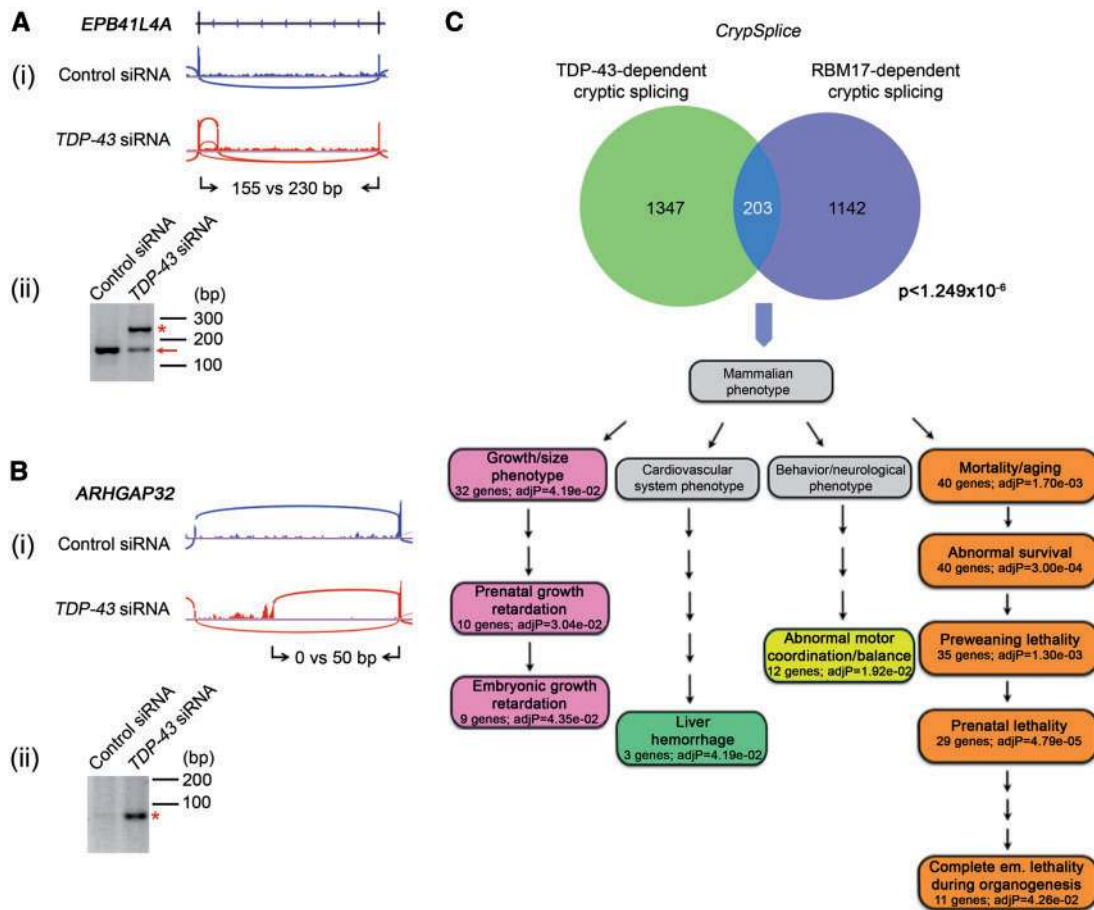
Figure 6. Cryptic splicing changes commonly regulated by RBM17 and TDP-43 are essential to cell survival. CrypSplice analyses reveal substantially more TDP-43-dependent cryptic splicing repression events than previously reported. Some of these predicted changes were validated in human HeLa cells with *TDP-43* knocked down using siRNA. The splicing validation PCRs of a previously reported hit, *EPB41L4A*, and a novel hit identified by CrypSplice, *ARHGAP32*, are shown in (**A**) and (**B**) respectively. Sashimi plots depicting the connection between reads are shown in (**i**). PCR gel images are shown in (**ii**) with the expected products in the controls (if present) indicated with red arrows and those in the knockouts indicated by asterisks. (**C**) Significant overlap is observed for genes with TDP-43- or RBM17-dependent cryptic splicing repression. Significance value was calculated using the hypergeometric test with 14 000 genes as the gene universe. The common set of 203 genes was further analyzed for phenotype enrichment. Phenotypes with significant enrichment are shown in bright colored boxes with bold fonts. The number of genes associated with each significantly enriched phenotype and their respective adjusted *P* values are also indicated.

the common targets of TDP-43 cryptic splicing repression in both species (Supplementary Material, Fig. S9). Interestingly, TDP-43-dependent cryptic splicing junctions in 148 genes were found in equivalent junctions between human and mouse genomes, further indicating that they are common targets of TDP-43 (Supplementary Material, Table S11). TDP-43 binds to UG repeats and UG repeats have been found close to TDP-43 associated cryptic exons (13). We further explored whether TDP-43 dependent cryptic junctions identified by CrypSplice were associated with UG repeats. We took cryptic junctions that were at least 10% incorporated (determined from the percentage of cryptic junctional reads in total reads from 5′exon) and searched for UG repeats 400 bp upstream and 100 bp downstream. We found that UG repeats exist in regions flanking most of the TDP-43-dependent cryptic splicing junctions, with majority of them having two repeats (Supplementary Material, Fig. S10).

Although a paralleled comparison between two different species is likely to reveal some of the common direct targets of TDP-43-dependent cryptic splicing modulation, we reasoned that a comparison between two RBPs sharing some common phenotypes (e.g. embryonic lethality, cell death upon RBP depletion) and splicing outcome (e.g. widespread cryptic splicing

dysregulation) might uncover converging molecular themes. To this end, we performed phenotype enrichment analyses on the 203 genes sharing cryptic splicing de-repression when either *Rbm17* or *Tdp-43* was depleted in mouse cells. We found significant enrichment for genes associated with abnormal survival (40 genes), prenatal lethality (29 genes) and complete embryonic lethality during organogenesis (11 genes) (Fig. 6C, Supplementary Material, Table S12). These genes could be key contributors to the embryonic lethality and cell death due to the loss of RBM17 or TDP-43.

Discussion

We have shown that RBM17 is crucial for cell survival and that its loss leads to widespread disruption in splicing, especially the splicing of cryptic junctions. Previous studies demonstrated that RBM17 affects alternative splicing of two genes, *Sxl* and *FAS* (18,20,21,25), but the splicing function of RBM17 had not been established at the whole transcriptome level. Knockdown of *RBM17* in cultured mammalian cells led to exon 6 inclusion in *FAS*, producing the pro-apoptotic form of the protein (18,25). In neurons lacking RBM17, we did not identify any alternative

splicing changes in *Fas* (Supplementary Material, Table S3 and S4). Upregulation of apoptotic pathways upon the loss of RBM17 in mutant neurons was thus not driven by the generation of pro-apoptotic FAS but by other mechanisms (discussed below). The observation that RBM17 regulates isoform splicing of *Sxl* and *FAS* raised the possibility that global regulation of isoform splicing could be a general function of RBM17. Our genome-wide splicing analyses, however, suggest that RBM17 regulates alternative isoform splicing for only relatively few (134) genes, and rather predominantly regulates the splicing of cryptic junctions. Interestingly, more than 90% of the genes with RBM17-dependent cryptic splicing changes gained additional junctions in the mutants, suggesting RBM17 normally represses the splicing of cryptic junctions and its loss leads to the inclusion of intronic elements in mature transcripts. How RBM17 executes such repression is unclear at this point. One possibility is that RBM17 and its interacting partners could block the utilization of these intronic elements or cryptic exons by directly binding to the cryptic junctions or adjacent regions. This scenario would be similar to what has been described for the interaction of SXL and RBM17 blocking the selection of *Sxl* exon 3 (20). We attempted to identify consensus motifs surrounding RBM17-dependent cryptic splice sites (Supplementary Material, Table S6), but further experiments with crosslinking and immunoprecipitation followed by deep sequencing (CLIPseq) will be necessary to identify RBM17 binding sites, and to address the mechanistic functions of RBM17 in cryptic splice site repression.

Exonization of intronic coding cassettes is predicted to create frameshifts or introduce PTCs in nearly 80% of the cases (35,36). Transcripts harboring PTCs are targeted for rapid degradation through a quality-control mechanism called nonsense-mediated decay (NMD), which requires the recruitment of up frameshift proteins and other NMD-activating components during inefficient translation termination at the PTC (37,38). In our study, of the 22 validated splicing changes, 16 were cryptic exon inclusion, of which 15 were predicted to produce a PTC (Supplementary Material, Table S5). We further tested whether five of these genes represent *bona fide* NMD targets in an *in vitro* cell assay system. Our results showed that four out of the five genes with RBM17-dependent cryptic exon-inclusion events generated isoforms that were sensitive to cycloheximide, an indirect inhibitor of NMD, suggesting that these isoforms were NMD targets (Supplementary Material, Fig. S11). In several validated cryptic exon inclusion events, cryptic exons were incorporated into >20% of the total transcripts (Supplementary Material, Table S5). If these were *bona fide* NMD targets, one would expect a concomitant decrease in gene expression. However, our analysis did not find a correlation between incorporation of cryptic events and downregulation in gene expression. We initially speculated that the TRAP approach, compared with the conventional total RNA extraction method, might have captured more NMD targets that were bound to the ribosomes during their pioneer rounds of translation before the recruitment of NMD-activating machinery. But we found this unlikely as the relative abundance of cryptic exon inclusion events was similar with or without the use of TRAP method (Supplementary Material, Fig. S12). The lack of global gene downregulation could be largely due to the fact that majority of the aberrantly spliced transcripts contribute to <20% of the transcripts. Therefore, even though the use of cryptic exons creates targets for NMD, it would only affect a small portion of transcripts and might not decrease gene expression significantly. In addition, the relatively small sample size (n = 3) in

our BacTRAP-RNAseq studies might also led to the reduced sensitivity in detecting small gene expression changes. Indeed, when we increased the sample size (n = 6) and re-analyzed the expression of four genes with high percentage (∼20%) incorporation of cryptic exons using qRT-PCR, we found that we could now detect a significant downregulation in one of the genes (Supplementary Material, Fig. S13). Nonetheless, our enrichment analyses suggest that genes with abnormal cryptic splicing changes are highly associated with the phenotypes observed in *Rbm17* knockouts, arguing for a strong contribution from these aberrant transcripts to the mutant phenotypes. We speculate that the abnormal transcripts in mutant cells lacking *Rbm17* can produce altered, in many cases truncated (Supplementary Material, Table S5), proteins with dominant-negative or toxic gain-of-function activities, thereby leading to cellular stress, upregulation of apoptotic pathways and ultimately cell death.

Both RBM17 and TDP-43 mouse knockouts die at an early embryonic stage before E8.5 (Supplementary Material, Table S1) (33,34), probably due to upregulation of apoptosis (Purkinje neurons lacking *Rbm17* upregulate apoptotic pathways (Fig. 3), while mouse embryonic stem cells depleted for *Tdp-43* undergo apoptosis (13)). At the molecular level, we found that both RBPs repress cryptic splicing for over a thousand genes. Comparing the genes displaying RBM17- and TDP-43-dependent cryptic splicing repression to identify potential common mechanisms underlying the apoptosis, we found a set of overlapping genes that are significantly associated with cell death and early lethality. Correcting the cryptic splicing defects in TDP-43 deficient cells can rescue apoptosis, providing a direct link between abnormal cryptic splicing repression and cell death (13). Since we have greatly expanded the list of genes with TDP-43-dependent cryptic splicing, it remains to be determined whether the rescue effect is true for all the targets or only a subset of genes.

Advances in high throughput sequencing technology, especially in RNA-seq, have made systematic investigation of alternative splicing possible, but the lack of suitable tools has made detection of novel cryptic splicing events challenging. The method developed in this study, CrypSplice, complements the existing methods that focus on gene isoforms and reference exons and should thus be an invaluable tool for the research community. In light of the two recent reports demonstrating that abnormal cryptic splicing activities by RBPs underlie disease pathogenesis and abnormal neuronal function (13,14), revisiting existing RNA-seq data using CrypSplice might yield novel biological insights that would have been missed by manual or heuristic search. In this study, re-analyzing the published TDP-43 data using CrypSplice markedly extended the original findings. Altogether, our study reveals a key *in vivo* function of RBM17 and suggests that extensive cryptic splicing repression mediated by RBM17 and TDP-43 are essential for cell survival. CrypSplice provides a new tool to take a fresh look at past RNA-seq datasets to uncover novel biological insights. This is particularly exciting given the amount of available RNA-seq data on disease-causing RBPs, for which cryptic splicing effects are completely unexplored for the majority of the proteins.

Overall our study demonstrates for the first time that the physiological function of RBM17 is crucial for cell survival and neuronal health, and that defects in RBM17 cause global splicing disruption and de-repression of cryptic splice sites. The list of genes sharing both RBM17- and TDP-43-dependent cryptic splicing de-repression provides an exciting opportunity to further dissect the molecular mechanism underlying neuronal death in the two neurodegenerative disease models.

Materials and Methods

Mouse handling

All procedures for mouse animal use were approved by the Institutional Animal Care and Use Committee for Baylor College of Medicine and Affiliates. Primers used for genotyping in the study are listed in Supplementary Material, Table S13.

Generation of RBM17 antibody

DNA sequences encoding 100–200 amino acid residues of RBM17 was cloned into *pGEX 4T-1* vector using *EcoRI* and *XhoI* sites. GST-RBM17 100-200 was then transformed into BL21 and was expressed and purified with GST antibody. This antigen was sent to Neuromab for the generation of anti-RBM17/SPF45 antibody (UCDavis/NIH NeuroMab facility Cat no. 73-234).

Immunohistochemistry and Purkinje cell pathology analysis

Mouse brains were dissected and immersed in 4% paraformaldehyde in PBS overnight and then prepared for paraffin embedding using standard protocol. The brain sections were cut to a thickness of 5 μm and were immunostained with anti-RBM17 antibody (Neuromab) or anti-calbindin antibody (Sigma) and imaged using an Axio Scan Z1 microscope. X-gal staining was done as previously described in (39).

Mouse behavioral studies

Rotarod analysis, open field assay, and dowel test were performed using 8-week-old mice as previously described in (40–43). Statistical analyses were performed using GraphPad Prism software.

BacTRAP profiling and data analyses, quantitative RT-PCR, and splicing RT-PCR analyses

BacTRAP profiling was performed as previously described (44) with the following modifications: One cerebellum was homogenized in 1.5 ml of homogenization buffer and 60 μg each of 19C8 and 19F7 was used for each IP. Detailed protocol and the subsequent bioinformatic analyses are described in Supplementary Material. Quantitative RT-PCR was performed using Perfecta SYBR Green FastMix (Quanta Biosciences). Relative fold change in gene expression was normalized to *Rps16* and/or *Hprt* and calculated using the $2^{\Delta\Delta CT}$ method (45). Splicing RT-PCR reactions were performed using KOD Hot Start Polymerase (EMD Millipore) or EconoTaq PLUS GREEN 2X Master Mixes (Lucigen) and standard PCR protocols. PCR products were run on MetaPhor agarose (Lonza) gels. Primers used for PCRs are listed in Supplementary Material, Table S9 and S13.

TDP-43 siRNA knockdown in human HeLa cells

HeLa cells grown in a 6-well plate were transfected with 40 nM of *TDP-43* siRNA (Sigma EHU109221) or control siRNA (Qiagen 1027280) in triplicates and cultured for 72 hours. Total RNA isolation was performed using TRIzol (Invitrogen) and Aurum Total RNA Fatty and Fibrous Tissue Kit (BioRad). 2.5 μg of total RNA was used for cDNA synthesis using the M-MLV RT Kit (Invitrogen). qPCRs were performed using the 2X GeneAmp Fast PCR Master Mix (BioRad) and PCRs were performed using the

KOD Hot Start DNA Polymerase (EMD Millipore). PCR products were analyzed using MetaPhor agarose gels (Lonza).

CrypSplice

Processing splice junctions. CrypSplice avoids remapping of reads and relies on junction counts from any efficient read-mapping algorithms. In this study, we used junction counts from TopHat v2.0.9 (30) as input to CrypSplice. TopHat is advantageous in controlling true negatives (46) and the adopted beta binomial modeling (31) of CrypSplice compliment this by controlling false positives. In agreement to the definition of cryptic splicing given in the results section *CrypSplice* first filters out all known/database reported splicing junctions from the observed junction (TopHat output). Errors in sequencing may result in ambiguous junction boundaries (47,48). To account any such errors CrypSplice collapses junctions with a reciprocal overlap of greater than or equal to an overlap cut-off (here 95%). Accepted and rejected reciprocal overlaps are illustrated in Supplementary Material, Fig. S14. Coverage of the new collapsed junction is computed as the average respective junctions. Since no alignment is perfect, junctions with coverage less than a minimal expression threshold (here 10) are ignored (49,50) to minimize any alignment noise. However, user can tailor select the expression noise threshold depending upon the sequencing depth. Expression score of every resultant junction is computed by $S_i = \frac{J_i}{U_i}$, where S_i is the expression score of ith junction, J_i is the number of reads spanning the junction i and U_i is the respective 5′ coverage. J_i can then be abstracted as junction counts and U_i as total sample counts for beta-binomial modelling. Beta binomial test is advantageous in modelling count data to other non-parametric methods such as the Mann–Whitney test or the Kruskal–Wallis test especially when the number of samples per condition is small (typically three for biological samples). It differentiates intra- and inter-sample variations and minimizes large artificial test statistics for the junctions in highly expressed genes. Intra-sample variation is modelled with a binomial distribution and inter sample variation is modelled by treating the binomial distribution parameter as a random variable following beta distribution. Model parameters in beta-binomial distribution are inferred by maximum likelihood estimation (MLE) procedure. Finally, likelihood ratio test is used to infer the significance of the group differences.

Beta-Binomial model. Let $J \in \mathbb{N}$ denote the number of junction read, $U \in \mathbb{N}$ the total number counts spanning this region and $\mathbb{N}$ the set of natural numbers. Assume J is distributed according to a binomial distribution with success probability $r \in [0, 1]$, $p(J|r\ U) = \binom{U}{J} r^J (1-r)^{U-J}$. To capture the variations between biological replicates, we model r through a beta distribution with $\alpha > 0$ and $\beta > 0$, $p(r|\alpha\ \beta) = \pi^{\alpha-1}(1-\pi)^{\{\beta-1\}} B(\alpha,\ \beta)^{-1}$, where $B(\cdot, \cdot)$ is the beta function. For numerical stability, we can parameterize the beta distribution to $\pi = \alpha(\alpha+\beta)^{-1}$, $\rho = (\alpha+\beta)^{-1}$, where π is the expectation of the r and ρ represents the dispersion. The log-likelihood of the observed data is given by

$$L = \sum_{i=1}^{N} \left[\sum_{x=0}^{J_i} \log(\pi + x\rho) + \sum_{x=0}^{U_i - J_i - 1} \log(1 - \pi + x\rho) - \sum_{x=0}^{U_i - 1} \log(1 + x\rho) \right].$$

Assuming there are G groups in an experiment, we let L_g be the maximal log-likelihood value for group $g = 1, \ldots, G$. We propose to test the homogeneity of the groups by likelihood ratio

test, where the log likelihood ratio statistics S is given by $2\left(-L_0 + \sum_g L_g\right)$. S is approximately χ^2 distribution with $2(G-1)$ degrees of freedom. The null hypothesis of this test is that the expectation and dispersion of the different groups are equal.

Every junction is multiple testing corrected using Benjamini-Hochberg procedure (51). Significant junctions with an adjusted *P* value < 0.01 are predicted as cryptic splicing changes. We further classified cryptic events into three groups based on minimum expression threshold (Supplementary Material, Fig. S5): junction gains (observed only in case samples), junction losses (observed only in control samples), and differential junctions, which may not meet the full definition of *cryptic*. Junctions spanning more than one gene are marked as overlaps.

Accession number

Data are available at Gene Expression Omnibus GSE79020.

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

Funding

References

1. Wang, Z. and Burge, C.B. (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA*, **14**, 802–813.
2. Matera, A.G. and Wang, Z. (2014) A day in the life of the spliceosome. *Nat. Rev. Mol. Cell. Biol.*, **15**, 108–121.
3. Graubert, T.A., Shen, D., Ding, L., Okeyo-Owuor, T., Lunn, C.L., Shao, J., Krysiak, K., Harris, C.C., Koboldt, D.C., Larson, D.E. *et al.* (2012) Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nat. Genet.*, **44**, 53–57.
4. Cazzola, M., Rossi, M. and Malcovati, L. (2013) Biologic and clinical significance of somatic mutations of SF3B1 in myeloid and lymphoid neoplasms. *Blood*, **121**, 260–269.
5. Vance, C., Rogelj, B., Hortobagyi, T., De Vos, K.J., Nishimura, A.L., Sreedharan, J., Hu, X., Smith, B., Ruddy, D., Wright, P. *et al.* (2009) Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science*, **323**, 1208–1211.
6. Sreedharan, J., Blair, I.P., Tripathi, V.B., Hu, X., Vance, C., Rogelj, B., Ackerley, S., Durnall, J.C., Williams, K.L., Buratti, E. *et al.* (2008) TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science*, **319**, 1668–1672.
7. Elden, A.C., Kim, H.J., Hart, M.P., Chen-Plotkin, A.S., Johnson, B.S., Fang, X., Armakola, M., Geser, F., Greene, R., Lu, M.M. *et al.* (2010) Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature*, **466**, 1069–1075.
8. Neumann, M., Sampathu, D.M., Kwong, L.K., Truax, A.C., Micsenyi, M.C., Chou, T.T., Bruce, J., Schuck, T., Grossman, M., Clark, C.M. *et al.* (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science*, **314**, 130–133.
9. Johnson, J.O., Pioro, E.P., Boehringer, A., Chia, R., Feit, H., Renton, A.E., Pliner, H.A., Abramzon, Y., Marangi, G., Winborn, B.J. *et al.* (2014) Mutations in the Matrin 3 gene cause familial amyotrophic lateral sclerosis. *Nat. Neurosci.*, **17**, 664–666.
10. Hooper, J.E. (2014) A survey of software for genome-wide discovery of differential splicing in RNA-Seq data. *Hum. Genomics*, **8**, 3.
11. Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. and Sharp, P.A. (1986) Splicing of messenger RNA precursors. *Annu. Rev. Biochem.*, **55**, 1119–1150.
12. Kapustin, Y., Chan, E., Sarkar, R., Wong, F., Vorechovsky, I., Winston, R.M., Tatusova, T. and Dibb, N.J. (2011) Cryptic splice sites and split genes. *Nucleic Acids Res.*, **39**, 5837–5844.
13. Ling, J.P., Pletnikova, O., Troncoso, J.C. and Wong, P.C. (2015) TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD. *Science*, **349**, 650–655.
14. Eom, T., Zhang, C., Wang, H., Lay, K., Fak, J., Noebels, J.L. and Darnell, R.B. (2013) NOVA-dependent regulation of cryptic NMD exons controls synaptic protein levels after seizure. *Elife*, **2**,
15. Lim, J., Crespo-Barreto, J., Jafar-Nejad, P., Bowman, A.B., Richman, R., Hill, D.E., Orr, H.T. and Zoghbi, H.Y. (2008) Opposing effects of polyglutamine expansion on native protein complexes contribute to SCA1. *Nature*, **452**, 713–718.
16. Perry, W.L., Shepard, R.L., Sampath, J., Yaden, B., Chin, W.W., Iversen, P.W., Jin, S., Lesoon, A., O'Brien, K.A., Peek, V.L. *et al.* (2005) Human Splicing Factor SPF45 (RBM17) Confers Broad Multidrug Resistance to Anticancer Drugs When Overexpressed— a Phenotype Partially Reversed By Selective Estrogen Receptor Modulators. *Cancer Res.*, **65**, 6593–6600.
17. Crisci, A., Raleff, F., Bagdiul, I., Raabe, M., Urlaub, H., Rain, J.C. and Krämer, A. (2015) Mammalian splicing factor SF1 interacts with SURP domains of U2 snRNP-associated proteins. *Nucleic Acids Res.*, **43**, 10456–10473.
18. Corsini, L., Bonnal, S., Basquin, J., Hothorn, M., Scheffzek, K., Valcarcel, J. and Sattler, M. (2007) U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. *Nat. Struct. Mol. Biol.*, **14**, 620–629.
19. Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A. and Mann, M. (1998) Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat. Genet.*, **20**, 46–50.
20. Lallena, MaJ., Chalmers, K.J., Llamazares, S., Lamond, A.I. and Valcárcel, J. (2002) Splicing Regulation at the Second Catalytic Step by Sex-lethal Involves 3′ Splice Site Recognition by SPF45. *Cell*, **109**, 285–296.

21. Al-Ayoubi, A.M., Zheng, H., Liu, Y., Bai, T. and Eblen, S.T. (2012) Mitogen-activated protein kinase phosphorylation of splicing factor 45 (SPF45) regulates SPF45 alternative splicing site utilization, proliferation, and cell adhesion. *Mol. Cell Biol.*, **32**, 2880–2893.

22. Kimmel, R.A., Turnbull, D.H., Blanquet, V., Wurst, W., Loomis, C.A. and Joyner, A.L. (2000) Two lineage boundaries coordinate vertebrate apical ectodermal ridge formation. *Genes Dev.*, **14**, 1377–1389.

23. Barski, J.J., Dethleffsen, K. and Meyer, M. (2000) Cre recombinase expression in cerebellar Purkinje cells. *Genesis*, **28**, 93–98.

24. Zhang, X.M., Ng, A.H., Tanner, J.A., Wu, W.T., Copeland, N.G., Jenkins, N.A. and Huang, J.D. (2004) Highly restricted expression of Cre recombinase in cerebellar Purkinje cells. *Genesis*, **40**, 45–51.

25. Liu, Y., Conaway, L., Rutherford Bethard, J., Al-Ayoubi, A.M., Thompson Bradley, A., Zheng, H., Weed, S.A. and Eblen, S.T. (2013) Phosphorylation of the alternative mRNA splicing factor 45 (SPF45) by Clk1 regulates its splice site utilization, cell migration and invasion. *Nucleic Acids Res.*, **41**, 4949–4962.

26. Doyle, J.P., Dougherty, J.D., Heiman, M., Schmidt, E.F., Stevens, T.R., Ma, G., Bupp, S., Shrestha, P., Shah, R.D., Doughty, M.L. *et al.* (2008) Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell*, **135**, 749–762.

27. Heiman, M., Schaefer, A., Gong, S., Peterson, J.D., Day, M., Ramsey, K.E., Suárez-Fariñas, M., Schwarz, C., Stephan, D.A., Surmeier, D.J. *et al.* (2008) A Translational profiling approach for the molecular characterization of CNS cell types. *Cell*, **135**, 738–748.

28. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L. and Pachter, L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.*, **7**, 562–578.

29. Anders, S., Reyes, A. and Huber, W. (2012) Detecting differential usage of exons from RNA-seq data. *Genome Res.*, **22**, 2008–2017.

30. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.*, **14**, R36.

31. Pham, T.V., Piersma, S.R., Warmoes, M. and Jimenez, C.R. (2010) On the beta-binomial model for analysis of spectral count data in label-free tandem mass spectrometry-based proteomics. *Bioinformatics*, **26**, 363–369.

32. Kabashi, E., Valdmanis, P.N., Dion, P., Spiegelman, D., McConkey, B.J., Vande Velde, C., Bouchard, J.P., Lacomblez, L., Pochigaeva, K., Salachas, F. *et al.* (2008) TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nat. Genet.*, **40**, 572–574.

33. Sephton, C.F., Good, S.K., Atkin, S., Dewey, C.M., Mayer, P., Herz, J. and Yu, G. (2010) TDP-43 is a developmentally regulated protein essential for early embryonic development. *J. Biol. Chem.*, **285**, 6826–6834.

34. Wu, L.S., Cheng, W.C., Hou, S.C., Yan, Y.T., Jiang, S.T. and Shen, C.K.J. (2010) TDP-43, a neuro-pathosignature factor, is essential for early mouse embryogenesis. *Genesis*, **48**, 56–62.

35. Sorek, R., Shamir, R. and Ast, G. (2004) How prevalent is functional alternative splicing in the human genome?. *Trends Genet.*, **20**, 68–71.

36. Sorek, R. (2007) The birth of new exons: mechanisms and evolutionary consequences. *RNA*, **13**, 1603–1608.

37. Lykke-Andersen, S. and Jensen, T.H. (2015) Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat. Rev. Mol. Cell. Biol.*, **16**, 665–677.

38. Popp, M.W. and Maquat, L.E. (2013) Organizing principles of mammalian nonsense-mediated mRNA decay. *Annu. Rev. Genet.*, **47**, 139–165.

39. Huang, W.H., Tupal, S., Huang, T.W., Ward, Christopher S., Neul, Jeffery L., Klisch, Tiemo J., Gray, Paul A. *et al.* (2012) Atoh1 governs the migration of postmitotic neurons that shape respiratory effectiveness at birth and chemoresponsiveness in adulthood. *Neuron*, **75**, 799–809.

40. Watase, K., Weeber, E.J., Xu, B., Antalffy, B., Yuva-Paylor, L., Hashimoto, K., Kano, M., Atkinson, R., Sun, Y., Armstrong, D.L. *et al.* (2002) A long CAG repeat in the mouse Sca1 locus replicates SCA1 features and reveals the impact of protein solubility on selective neurodegeneration. *Neuron*, **34**, 905–919.

41. Fryer, J.D., Yu, P., Kang, H., Mandel-Brehm, C., Carter, A.N., Crespo-Barreto, J., Gao, Y., Flora, A., Shaw, C., Orr, H.T. *et al.* (2011) Exercise and genetic rescue of SCA1 via the transcriptional repressor capicua. *Science*, **334**, 690–693.

42. Jafar-Nejad, P., Ward, C.S., Richman, R., Orr, H.T. and Zoghbi, H.Y. (2011) Regional rescue of spinocerebellar ataxia type 1 phenotypes by 14-3-3ϵ haploinsufficiency in mice underscores complex pathogenicity in neurodegeneration. *Proc. Natl. Acad. Sci. USA*, **108**, 2142–2147.

43. Park, J., Al-Ramahi, I., Tan, Q., Mollema, N., Diaz-Garcia, J.R., Gallego-Flores, T., Lu, H.C., Lagalwar, S., Duvick, L., Kang, H. *et al.* (2013) RAS-MAPK-MSK1 pathway modulates ataxin 1 protein levels and toxicity in SCA1. *Nature*, **498**, 325–331.

44. Heiman, M., Kulicke, R., Fenster, R.J., Greengard, P. and Heintz, N. (2014) Cell type–specific mRNA purification by translating ribosome affinity purification (TRAP). *Nat. Protocols*, **9**, 1282–1291.

45. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2-\Delta\Delta CT$ Method. *Methods*, **25**, 402–408.

46. Kumar, P.K., Hoang, T.V., Robinson, M.L., Tsonis, P.A. and Liang, C. (2015) CADBURE: A generic tool to evaluate the performance of spliced aligners on RNA-Seq data. *Sci. Rep.*, **5**, 13443.

47. Burset, M., Seledtsov, I.A. and Solovyev, V.V. (2001) SpliceDB: database of canonical and non-canonical mammalian splice sites. *Nucleic Acids Res.*, **29**, 255–259.

48. Richterich, P. (1998) Estimation of errors in "raw" DNA sequences: a validation study. *Genome Res.*, **8**, 251–259.

49. Xiong, H.Y., Alipanahi, B., Lee, L.J., Bretschneider, H., Merico, D., Yuen, R.K., Hua, Y., Gueroussov, S., Najafabadi, H.S., Hughes, T.R. *et al.* (2015) RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science*, **347**, 1254806.

50. Yalamanchili, H.K., Li, Z., Wang, P., Wong, M.P., Yao, J,,. and Wang, J. (2014) SpliceNet: recovering splicing isoform-specific differential gene networks from RNA-Seq data of normal and diseased samples. *Nucleic Acids Res.*, **42**, e121.

51. Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N. and Golani, I. (2001) Controlling the false discovery rate in behavior genetics research. *Behav. Brain Res.*, **125**, 279–284.