Cell Reports
Resource

HITS-CLIP and Integrative Modeling Define the Rbfox Splicing-Regulatory Network Linked to Brain Development and Autism

Sebastien M. Weyn-Vanhentenryck,[1,9] Aldo Mele,[2,9] Qinghong Yan,[1,9] Shuying Sun,[3,4] Natalie Farny,[5] Zuo Zhang,[3,6] Chenghai Xue,[3] Margaret Herre,[2] Pamela A. Silver,[5] Michael Q. Zhang,[7,8] Adrian R. Krainer,[3] Robert B. Darnell,[2,*] and Chaolin Zhang[1,*]

[1]Department of Systems Biology, Department of Biochemistry and Molecular Biophysics, Center for Motor Neuron Biology and Disease, Columbia University, New York, NY 10032, USA
[2]Howard Hughes Medical Institute, Laboratory of Molecular Neuro-Oncology, Rockefeller University, New York, NY 10065, USA
[3]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA
[4]Ludwig Institute for Cancer Research, University of California, San Diego, La Jolla, CA 92093, USA
[5]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA
[6]Merck Research Laboratories, Merck & Co., Inc., Rahway, NJ 07065, USA
[7]Department of Molecular and Cell Biology, Center for Systems Biology, The University of Texas at Dallas, Richardson, TX 75080, USA
[8]Bioinformatics Division, Center for Synthetic and Systems Biology, TNLIST, Tsinghua University, Beijing 100084, China
[9]These authors contributed equally to this work
*Correspondence: darnelr@rockefeller.edu (R.B.D.), cz2294@columbia.edu (C.Z.)
http://dx.doi.org/10.1016/j.celrep.2014.02.005

SUMMARY

The RNA binding proteins Rbfox1/2/3 regulate alternative splicing in the nervous system, and disruption of Rbfox1 has been implicated in autism. However, comprehensive identification of functional Rbfox targets has been challenging. Here, we perform HITS-CLIP for all three Rbfox family members in order to globally map, at a single-nucleotide resolution, their in vivo RNA interaction sites in the mouse brain. We find that the two guanines in the Rbfox binding motif UGCAUG are critical for protein-RNA interactions and crosslinking. Using integrative modeling, these interaction sites, combined with additional datasets, define 1,059 direct Rbfox target alternative splicing events. Over half of the quantifiable targets show dynamic changes during brain development. Of particular interest are 111 events from 48 candidate autism-susceptibility genes, including syndromic autism genes *Shank3*, *Cacna1c*, and *Tsc2*. Alteration of Rbfox targets in some autistic brains is correlated with downregulation of all three Rbfox proteins, supporting the potential clinical relevance of the splicing-regulatory network.

INTRODUCTION

The Rbfox proteins are a family of neuron- and muscle/heart-specific RNA binding proteins (RBPs) encoded by three genes—*Rbfox1* (*Fox-1* or *A2bp1*), *Rbfox2* (*Fox-2* or *Rbm9*), and *Rbfox3* (*Fox-3*, *Hrnbp3*, or *NeuN*)—that are conserved in vertebrates, flies, and worms. Rbfox1 and Rbfox2 are exclusively or preferentially expressed in neurons, heart, and muscles, whereas Rbfox3 is specifically expressed in postmitotic neurons. In humans, chromosomal translocation or copy number variation affecting *RBFOX1* has been found in patients with several neurological disorders, including epilepsy, intellectual disability (Bhalla et al., 2004), schizophrenia (Xu et al., 2008), and autism (Martin et al., 2007; Sebat et al., 2007).

At the molecular level, Rbfox proteins are known as tissue-specific splicing factors that bind to the (U)GCAUG element frequently conserved across vertebrate species (Jin et al., 2003; Minovitsky et al., 2005; Ponthier et al., 2006; Underwood et al., 2005). We previously performed genome-wide bioinformatic prediction of Rbfox target exons based on phylogenetically conserved motif sites (Zhang et al., 2008), leading to the identification of >1,000 alternative or constitutive exons that are potentially regulated by Rbfox, many found within transcripts encoding proteins important for neuromuscular functions. Characterization of the splicing pattern of the Rbfox targets revealed a position-dependent RNA map predictive of Rbfox action. According to this map, Rbfox binding in the downstream intron activates exon inclusion and binding in the alternative exon or upstream intron represses exon inclusion, consistent with observations from several tissue-specific exons (Jin et al., 2003; Underwood et al., 2005). Such a map was previously found for another neuron-specific splicing factor Nova and is now recognized as a more general rule of alternative splicing regulation (Licatalosi et al., 2008; Ule et al., 2006).

Despite recent progress (Barash et al., 2010; Ray et al., 2013; Ule et al., 2006; Zhang et al., 2013), the small sizes of RBP binding motifs limit the ability of motif-based bioinformatic target prediction to achieve both high specificity and sensitivity. To map in vivo protein-RNA interaction sites on a genome-wide scale, crosslinking and immunoprecipitation followed by high-throughput sequencing (HITS-CLIP) have been developed to

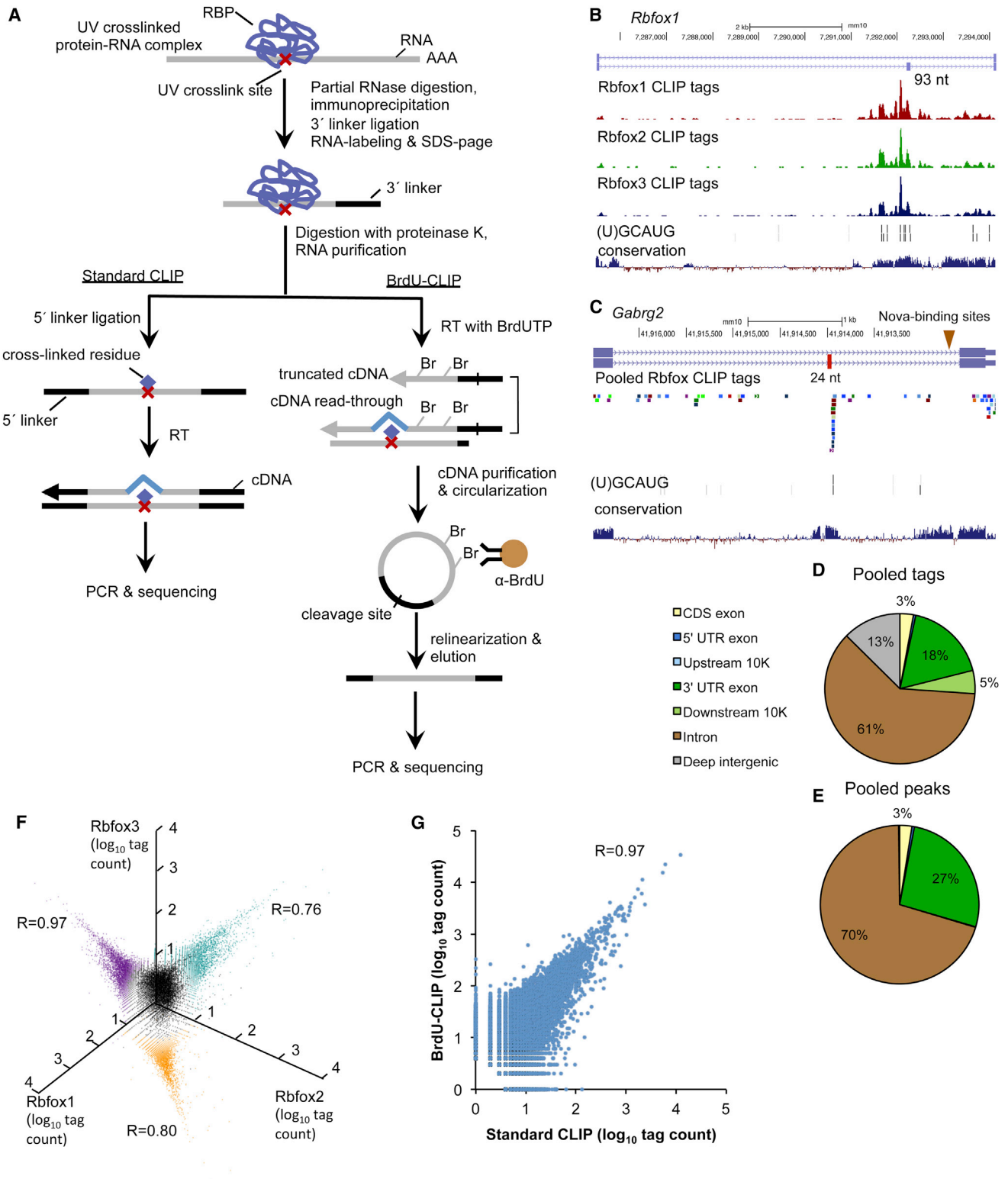

Figure 1. HITS-CLIP Maps Rbfox-RNA Interaction Sites in Mouse Brain on a Genome-wide Scale

(A) A schematic illustration of two HITS-CLIP protocols used to map Rbfox binding sites.

(B) A UCSC Genome Browser view of Rbfox1, 2, and 3 CLIP data in an alternatively spliced region of a 93 nt cassette exon in *Rbfox1* is shown. Rbfox1, 2, and 3 CLIP data are shown in separate wiggle tracks above the coordinates of UGCAUG and GCAUG elements and the phyloP conservation score.

(legend continued on next page)

isolate RNA fragments directly bound by an RBP of interest (Darnell, 2010; Licatalosi et al., 2008; Moore et al., 2014; Ule et al., 2005a). HITS-CLIP has been used to map the Rbfox2 binding sites in thousands of genes in human embryonic stem cells, including those important for splicing regulation as predicted by the RNA map (Yeo et al., 2009).

To understand the physiological function of Rbfox proteins in the mammalian brain, knockout (KO) mouse models have been generated. CNS-specific depletion of Rbfox1 results in an increased susceptibility of mice to seizures and in overexcitability of neurons in the dentate gyrus (Gehman et al., 2011). CNS depletion of Rbfox2 results in defects in cerebellar development (Gehman et al., 2012). Comparison of wild-type (WT) versus Rbfox1 or Rbfox2 KO brains using exon-junction microarrays has identified multiple Rbfox-dependent exons (Gehman et al., 2011, 2012). However, the number of exons identified using this approach is quite small (20 and 29 exons, respectively), compared to the number of Rbfox binding sites determined by bioinformatic prediction or CLIP data, presumably due to compensatory upregulation of Rbfox2 in *Rbfox1* KO mice and vice versa. Given that different Rbfox family members have highly similar protein sequences, especially in their RNA-recognition motif (RRM)-type RNA binding domain (RBD; $\geq$94% amino acid identity), they are expected to bind and regulate largely overlapping sets of transcripts (Gallagher et al., 2011; Gehman et al., 2012).

Until now, a comprehensive and accurate target splicing-regulatory network of the Rbfox proteins has not been defined, due in part to the lack of a genome-wide high-resolution map of the Rbfox interaction sites in the brain and to the lack of effective computational methods to couple protein-RNA interactions with splicing changes as a means of identifying direct, functional targets. Here, we used HITS-CLIP to globally map the RNA interaction sites of all three Rbfox family members and complemented the CLIP data with RNA sequencing (RNA-seq) data to identify exons responsive to perturbation of Rbfox. Importantly, we probabilistically weighed and combined these and additional data sets to define the functional target transcripts directly regulated by Rbfox using an integrative modeling approach (Zhang et al., 2010). The resulting network allowed us to reveal the role of Rbfox proteins in regulating global dynamic splicing changes during brain development and highlight promising downstream targets implicated in autism.

RESULTS

Rbfox1, 2, and 3 HITS-CLIP in Mouse Brain

Considering the possibility that each Rbfox family member might differ in binding specificity despite their apparent functional redundancy, we performed HITS-CLIP experiments for all members of the family individually using mouse whole-brain tissue. We first confirmed that the antibodies we used did not cross-react with different members and that they efficiently immunoprecipitated (IP) the targeted protein with minimal background under standard CLIP conditions (Figure S1; Supplemental Notes). Next, we used two different strategies to clone and amplify the isolated RNA fragments (Figure 1A). The first protocol, denoted as standard CLIP, was performed as described previously (Darnell, 2010; Licatalosi et al., 2008; Moore et al., 2014; Ule et al., 2005a). In this protocol, RNA linkers are ligated to the 5′ and 3′ ends of the RNA fragments (Figure 1A, left branch) and are later used for RT-PCR amplification. We and several other groups have previously noted that after proteinase K digestion of the crosslinked protein-RNA complex, one or a few amino acids might remain attached to the RNA at the crosslink site, which causes informative errors at the crosslink site during reverse transcription (Granneman et al., 2009; Ule et al., 2005a). These crosslinking-induced mutation sites (CIMS) provide a footprint of protein-RNA crosslinking and can be leveraged to determine protein-RNA interactions at a single nucleotide resolution (Moore et al., 2014; Zhang and Darnell, 2011). However, reverse transcription can abort prematurely at these sites, resulting in truncated cDNAs that lack the 5′ adaptor required for PCR (König et al., 2010; Sugimoto et al., 2012). To capture both truncated and nontruncated cDNAs, we developed a second CLIP protocol named bromodeoxyuridine (BrdU)-CLIP (Figure 1A, right branch). This protocol bears some conceptual similarity to individual nucleotide resolution CLIP or iCLIP (König et al., 2010). After ligation of the 3′ linker, purified RNA is reverse transcribed to introduce 5′ and 3′ PCR adaptor sequences separated by an apurinic/apyrimidinic endonuclease (APE) cleavage site. This is followed by the circularization of both readthrough and truncated cDNAs and relinearization of cDNA via the cleavage site to place the 5′ and 3′ adaptor sequences in the correct orientation. One key difference between BrdU-CLIP and iCLIP is the incorporation of BrdUTP into the cDNA during reverse transcription so that the resulting cDNA can be purified in a stringent manner using an antibody that specifically recognizes BrdU (Core et al., 2008; Ingolia et al., 2009).

To evaluate the robustness of the Rbfox interaction sites, we prepared HITS-CLIP libraries for Rbfox1, Rbfox2, and Rbfox3 with four, four, and five biological replicates, respectively, which together resulted in about 870 million raw reads (CLIP tags). After stringent filtering, processing, and mapping (Moore et al., 2014; Zhang et al., 2010) (Experimental Procedures), we obtained a total of 4.6 million unique CLIP tags that represent independent captures of protein-RNA interactions, including 1,460,387 tags for Rbfox1, 868,366 tags for Rbfox2, and 2,308,632 tags for

(C) Similar to (B), except that the alternatively spliced region of *Gabrg2* exon 9 is shown. CLIP data of Rbfox1, 2, and 3 are pooled together and shown in a single track with different colors representing the CLIP tags obtained in independent CLIP experiments. The position of Nova binding is indicated by the arrowhead (top right).
(D and E) Genomic distribution of Rbfox1, 2, and 3 CLIP tags pooled together (D), and the resulting genic CLIP tag cluster peaks (E) are shown. Due to incompleteness of 5′ and 3′ UTR annotations, each gene is extended for 10 kb in both directions; these regions are listed as separate categories.
(F) Pairwise correlation of CLIP data among Rbfox proteins based on the number of CLIP tags per cluster. Each cluster is represented as a black dot positioned in three-dimensional (3D) space. Comparisons between each pair of proteins are shown in 2D planes, obtained by projecting the black dots into their respective 2D space (colored dots). Pearson correlation of each pairwise comparison is indicated.
(G) Correlation of CLIP tags derived from the standard and BrdU-CLIP protocols, based on the number of CLIP tags per cluster.

Rbfox3. Between 59% and 65% of these CLIP tags are located in introns, consistent with the known role of Rbfox proteins in regulating alternative splicing; an additional 23%–28% unique CLIP tags are located in exons, mostly in the 3′ UTRs.

Rbfox1, 2, and 3 Have Similar Protein-RNA Interaction Profiles

Initial inspection of the CLIP tag distribution suggests that the interaction profiles of the three Rbfox family members are very similar. For example, the *Rbfox1* transcripts contain a cassette exon of 93 nt (Figure 1B) encoding part of the RRM of the protein. Its skipping as a result of autoregulation generates a dominant-negative form that lacks RNA binding capability (Baraniak et al., 2006; Damianov and Black, 2010). Our CLIP data show that all three Rbfox family members bind to the upstream intronic sequences harboring a cluster of conserved UGCAUG elements, suggesting that this exon is under both auto- and cross-regulation by all family members. We also previously demonstrated that GABA receptor gamma 2 subunit (*Gabrg2*) exon 9 is under the synergistic regulation of Rbfox and Nova when they bind near the 5′ and 3′ splice sites of the downstream intron, respectively, based on detailed mutation analysis and splicing reporter assays in cell culture (Dredge and Darnell, 2003; Zhang et al., 2010). Our CLIP data now confirmed that Rbfox proteins indeed bind to the expected site in vivo in the brain (Figure 1C).

To quantitatively compare the RNA binding profiles of different Rbfox family members, we defined a nonredundant set of Rbfox-RNA interaction sites using all unique CLIP tags pooled together (Figure 1D; Experimental Procedures). A stringent set of 41,182 genic CLIP tag clusters with at least one statistically significant peak ($p < 0.01$) was obtained (Table S1), 70% of which are located in introns and the other 30% are in exons (Figure 1E). Then, we counted the number of CLIP tags per cluster for each protein. CLIP tags for different members are very well correlated in each pairwise comparison, especially between Rbfox1 and Rbfox3 (Pearson correlation $R = 0.97$); the correlation between Rbfox2 and the other two members is somewhat lower ($R = 0.76$–0.80; Figure 1F). In addition, we confirmed that the two CLIP protocols gave very reproducible results in the global profiles ($R = 0.97$; Figure 1G). Based on these observations, we conclude that the three Rbfox family members have similar RNA-interaction profiles on a genome-wide scale, consistent with the notion that their binding specificity is largely determined by their very similar RRMs. Although it remains possible that a small proportion of the binding sites could be preferentially recognized by a specific member, for this work, CLIP tags of all three members were pooled together for further analysis.

A Single-Nucleotide Resolution Map of Rbfox Binding Sites by CIMS and CITS Analysis

Using two CLIP protocols in parallel allowed us to employ different strategies to pinpoint the exact Rbfox-RNA crosslink and interaction sites (Figures 2 and S2). For CLIP tags obtained by the standard protocol, we performed CIMS analysis using our established method (Figure 2A) (Moore et al., 2014; Zhang and Darnell, 2011). Nucleotide deletions were observed in 14% of standard CLIP tags, from which 1,424 reproducible CIMS were identified (false discovery rate [FDR] < 0.001). A substantial enrichment of the Rbfox binding motif GCAUG was observed in the immediate vicinity of the reproducible deletion sites (Figure 2B). In contrast, when we analyzed substitutions and insertions using the same method, we did not observe elevated motif enrichment near the mutation sites (data not shown), suggesting that crosslinking predominantly, if not exclusively, introduces deletions rather than insertions or substitutions in Rbfox CLIP.

We then examined the enrichment of UGCAUG or VGCAUG (V = non-U) relative to the crosslink sites with reproducible deletions in more detail (Figure 2B inset). UGCAUG is enriched 28- to 41-fold at positions −5, −4, and −1 relative to the crosslink site, corresponding to crosslinking at G2, U5, and G6 of the UGCAUG element. Interestingly, enrichment of VGCAUG is most predominant at position −1 relative to the crosslink site (64-fold), corresponding to crosslinking of G2 (Figure 2C). We also examined the base composition of the sequences around CIMS regardless of the presence of (U)GCAUG and observed a slight bias toward uridine compared to the flanking sequences (Figure S2A; see Discussion below). De novo motif analysis using sequences [-10,10] around CIMS uncovered (U)GCAUG as the only motif with strong enrichment (36% of 1,158 nonrepetitive CIMS in [−10,10], $E < 3.7 \times 10^{-320}$; Figure S2B). The first position of the motif is the most variable, which is consistent with previous findings that Rbfox binds to both UGCAUG and VGCAUG with high affinity (Jin et al., 2003; Ponthier et al., 2006). Additional deviations from the consensus appear to be tolerated to some extent (e.g., in positions 3 and 4), providing a partial explanation for why (U)GCAUG is not present at all crosslink sites (Figures S2E–S2G). Finally, we observed deletions in 6.2% of BrdU-CLIP tags, and analysis combining standard and BrdU-CLIP tags defined 2,298 CIMS (FDR < 0.001; Table S2).

The sensitivity of crosslink site identification by CIMS analysis is limited by the relatively low deletion rate among tags that are read through. We therefore looked for reproducible crosslinking induced truncation sites (CITS) in BrdU-CLIP data (Figure 2D; Experimental Procedures). Overall, 6,606 robust CITS were identified ($p < 0.001$; Table S3). Among these, the UGCAUG element is enriched 319-fold at a single position (−5) relative to CITS, corresponding to predominant crosslinking at G6 of the motif (Figure 2E). The same position was crosslinked in the VGCAUG element, although there is less enrichment of the motif (88-fold; Figure 2F). Analysis of the base composition [−10,10] around CITS revealed the UGCAUG motif directly (Figure S2C), and this was confirmed by de novo motif analysis (60% of 1000 randomly sampled nonrepetitive sites; $E < 1.9 \times 10^{-631}$; Figure S2D). As a control, we repeated the same analysis in the standard CLIP data, which presumably lacked truncated tags, and did not observe enrichment of the (U)GCAUG motif in these specific positions (Figures S2H and S2I).

Together, our data suggest that the two guanines G2 and G6 in the (U)GCAUG motif are particularly prone to crosslinking with the Rbfox protein. Consistent with this finding, examination of a previously determined NMR structure of the Rbfox1 RRM in complex with UGCAUGU RNA revealed that these two guanines are buried in two pockets of the RRM and are stabilized by multiple hydrogen bonds and stacking interactions (Figure 2G); mutations in each of these two nucleotides resulted in the largest increases in the free energy of binding (Auweter et al., 2006).

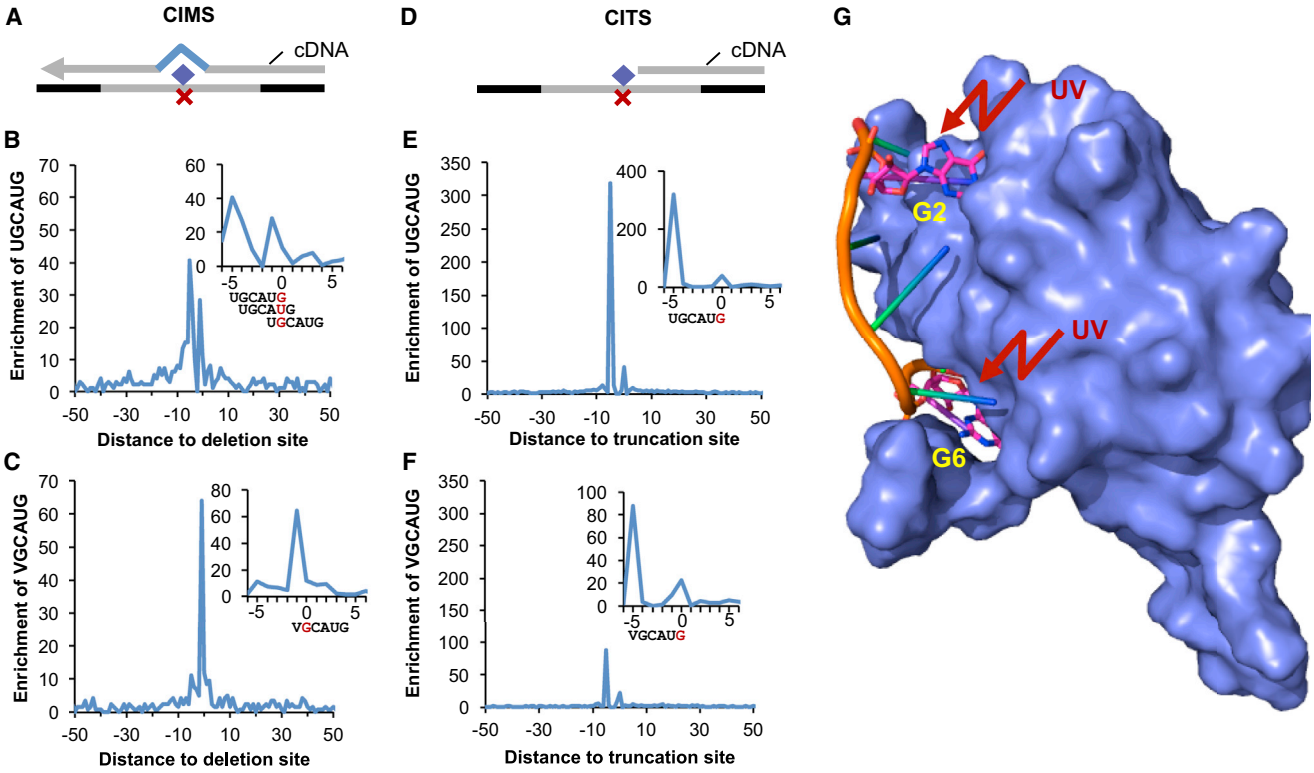

Figure 2. CIMS and CITS Analysis to Map Rbfox-RNA Interactions at a Single-Nucleotide Resolution

(A)–(C) and (D)–(F) are for crosslinking-induced mutation sites (CIMS) and crosslinking-induced truncation sites (CITS) analysis, respectively.

(A and D) A schematic illustration of CIMS (A) and CITS (D) is shown.

(B and E) Enrichment of UGCAUG around CIMS (deletions, B) and CITS (truncations, E) is calculated from the frequency of UGCAUG starting at each position relative to the inferred crosslink sites, normalized by the frequency of the element in flanking sequences. The inset shows a zoomed-in view, with the most frequent crosslink sites in the motif highlighted in red.

(C and F) Similar to (B) and (E), except that the enrichment of VGCAUG (V = non-U) around CIMS (C) and CITS (F) is shown.

(G) NMR structure of RbFox1 RRM (surface, pale blue) in complex with the UGCAUGU heptanucleotide (cartoon, rainbow; Protein Data Bank ID code 2ERR; Auweter et al., 2006). Highlighted are the two guanines G2 and G6 (pink) with predominant crosslinking.

Based on the single-nucleotide-resolution map of in vivo Rbfox interaction sites and on the characterized specificity of the proteins, we developed the motif enrichment and conservation score (MECS) of (U)GCAUG elements by comparing CLIP tag clusters and regions without CLIP tags (Figures S2J and S2K; Supplemental Results). A motif site with higher conservation receives a higher score, especially if it is located in an intronic region. A UGCAUG element receives a higher score than a VGCAUG element with the same level of conservation, reflecting greater enrichment of the former in CLIP tag clusters.

Identifying Rbfox-Dependent Exons Using RNA-Seq

We previously used HeLa cells with perturbed Rbfox1 or Rbfox2 expression to validate over half (55%–59%) of bioinformatically predicted Rbfox target alternative exons tested with RT-PCR (Zhang et al., 2008). We therefore used this established experimental system to expand the list of Rbfox-dependent exons by RNA-seq, which provides information complementary to the CLIP data. As we described previously (Zhang et al., 2008), HeLa cells were treated with a short hairpin RNA (shRNA) targeting Rbfox2 (shRbfox2) to generate stable knockdown (KD) of the protein, which is endogenously expressed at a level that is low but sufficient for splicing regulation; HeLa cells expressing the empty vector were used for comparison (control; Figure 3A, left panel). In total, RNA-seq of the control and shRbfox2 samples resulted in 60 million and 48 million paired-end reads, respectively, of which 62%–65% were mapped unambiguously to the genome or to the exon-junction database. Examination of the gene-expression level confirmed that Rbfox2 was specifically knocked down 3.3-fold (Figure 3A, right panel), consistent with the protein level changes observed from immunoblot analysis. Accordingly, we were able to identify 126 cassette exons, 17 tandem cassette exon events, and four mutually exclusive exon events showing Rbfox2-dependent inclusion or exclusion (FDR $\leq$ 0.1 and proportional change of exon inclusion $|\Delta I| \geq 0.1$) (Ule et al., 2005b) (Figure 3B; Table S4). Among the 22 cases of alternative exons we tested by RT-PCR (which includes three cases with a read coverage slightly below the threshold we used), 21 showed Rbfox2-dependent splicing (Figures 3C and 3D; Tables S4 and S5), giving a validation rate of 95%. The direction of Rbfox2-dependent splicing of these exons can be predicted by the position-dependent RNA map based on

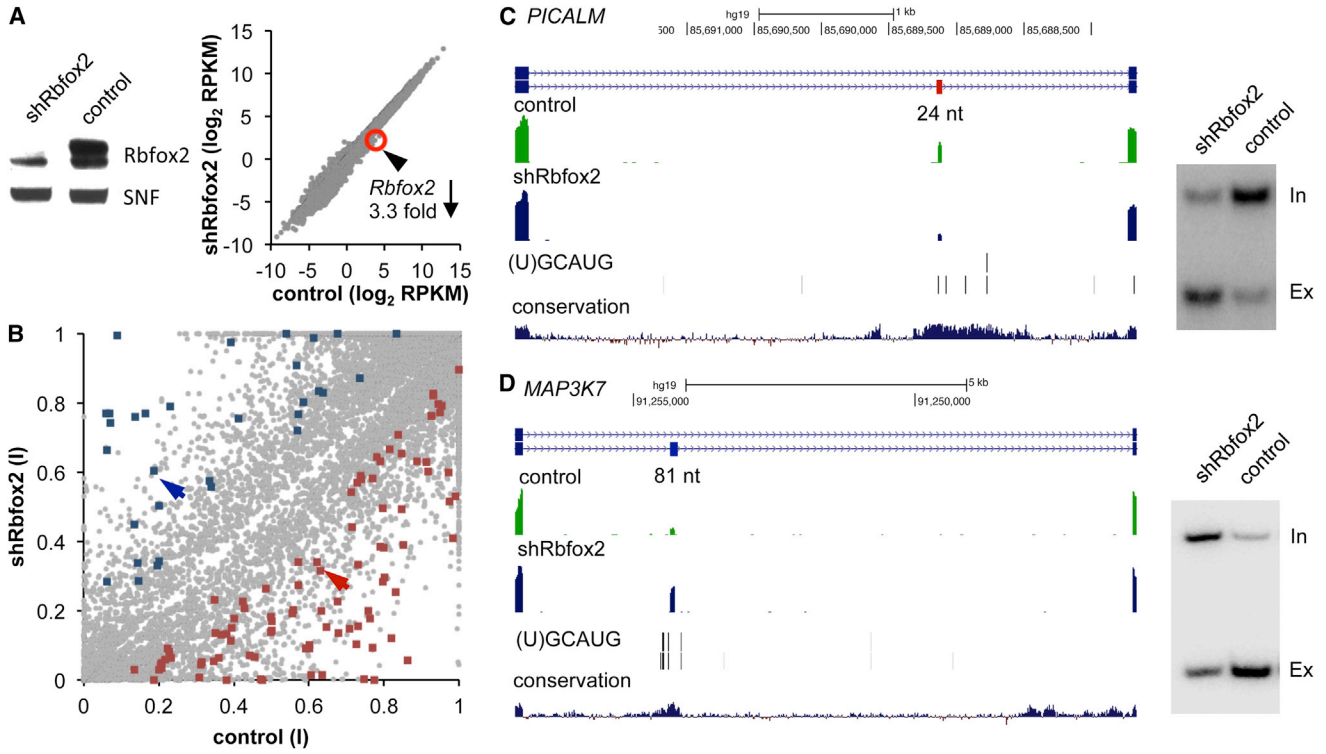

Figure 3. Identification of Rbfox2-Dependent Exons using RNA-Seq

(A) Left panel: Rbfox2 protein expression in HeLa cells with stable knockdown of Rbfox2 using a specific shRNA (shRbfox2) compared with cells treated with the vector (control), as evaluated by immunoblot analysis. The nucleosome remodeling complex protein SNF is used as a loading control. Right panel: the gene expression profiles in HeLa cells treated with shRbfox2 or control vectors are quantified by reads per kb/million (RPKM) using RNA-seq data. Rbfox2 is highlighted by the red circle.

(B) Proportional inclusion (I) of cassette exons in shRbfox and control HeLa cells. Exons with reduced and increased inclusion (FDR < 0.1 and $|\Delta I| > 0.1$; Ule et al., 2005b) in shRbfox2 compared with control HeLa cells are highlighted in red and blue, respectively.

(C and D) UCSC Genome Browser views of two examples of Rbfox-dependent exon inclusion (*PICALM*, C) or exclusion (*MAP3K7*, D), as indicted by the arrowheads in (B). Below the RNA-seq data are (U)GCAUG elements and phyloP conservation scores. The result of RT-PCR validation is shown on the right.

either the CLIP data derived from mouse brain or the bioinformatically predicted motif sites (data not shown), indicating that these Rbfox2-dependent exons are enriched in direct Rbfox targets in the brain.

Integrative Modeling Defines the Rbfox Target Splicing-Regulatory Network

To comprehensively define the functional target network directly regulated by the Rbfox proteins, we took an integrative modeling approach, which we have recently developed, and successfully applied it to study the Nova target network (Zhang et al., 2010). This method uses a Bayesian network to probabilistically weigh and combine multiple types of data complementary to each other: bioinformatically predicted motif sites represented by MECS scores, protein-RNA interaction sites mapped by Rbfox HITS-CLIP, Rbfox1-dependent splicing identified by comparison of WT with *Rbfox1* KO mouse brain using exon-junction microarrays (Gehman et al., 2011), Rbfox2-dependent splicing in HeLa cells as described above, tissue-specific splicing as measured by RNA-seq (only for training; Brawand et al., 2011), and evolutionary signatures including preservation of reading frame and conservation of alternative splicing pattern (Figure 4A; Experimental Procedures).

Focusing initially on cassette exons, we found that the estimated model parameters confirmed, quantified, and extended our understanding of Rbfox splicing regulation (Figures 4B–4E and S3A–S3E; Supplemental Results). For example, stronger motif sites are more likely to be bound by the protein (Figure 4B), and regions inferred to be bound by Rbfox have more CLIP tags than those inferred not to be bound (Figure 4C). In addition, the model was able to quantify the position-dependent RNA map: binding of Rbfox in the downstream intron is predicted to result in Rbfox-dependent inclusion with a probability of 0.99, whereas binding of Rbfox in the upstream intron or exon is predicted to result in repression with a probability of 0.75 and 0.61, respectively. Binding of Rbfox in both exon and upstream intron is expected to increase the probability of repression to 0.84 (Figure 4D).

The model was then applied to each annotated cassette exon in the mouse genome to predict the probability of its activation or repression by Rbfox through direct protein-RNA interactions. After using 10-fold cross-validation to ensure the model was

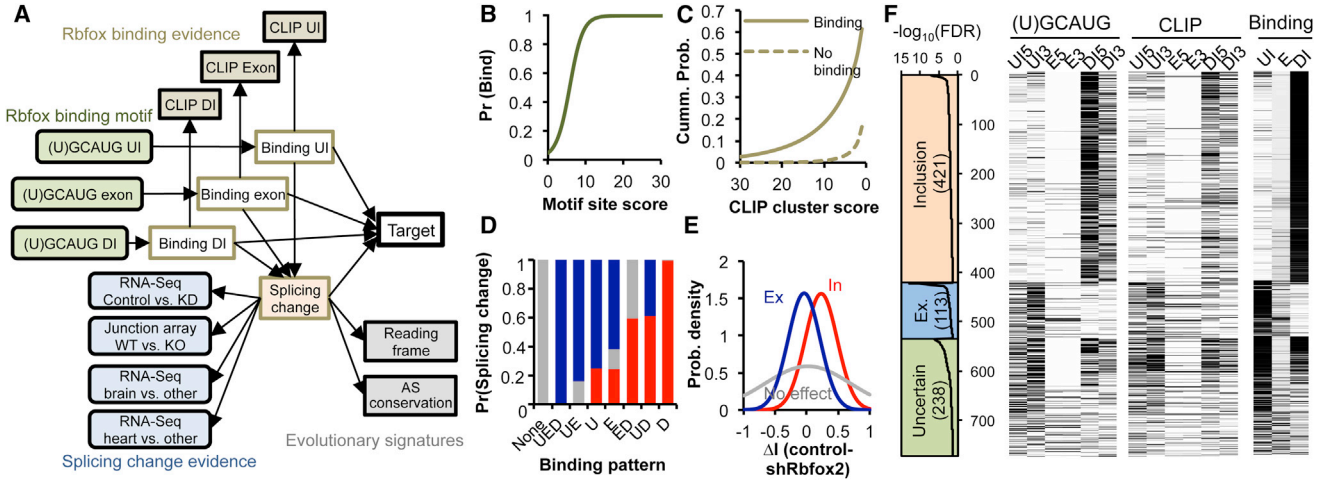

Figure 4. Integrative Modeling Predicts Rbfox Target Exons using a Bayesian Network

The model is trained using cassette exons.

(A) Design of the Bayesian network (BN). The 17 nodes (variables) model four types of data, including (U)GCAUG elements and CLIP tag clusters in each cassette exon or flanking upstream (UI) and downstream introns (DI), splicing change of exons with Rbfox depletion or among different tissues, and evolutionary signatures.

(B) The probability of Rbfox binding to regions with varying motif scores.

(C) The cumulative probability of CLIP tag cluster scores across all regions with or without inferred Rbfox binding.

(D) The probability of exons showing Rbfox-dependent inclusion (red), exclusion (blue), or no effect (gray), given the indicated combinatorial Rbfox binding patterns in the exon (E), upstream (U), and downstream (D) introns.

(E) The distribution of proportional splicing changes (ΔI) in Rbfox knockdown versus the control as measured by RNA-seq for exons with inferred Rbfox-dependent inclusion, exclusion, or without Rbfox regulation.

(F) Rbfox binding pattern for exons predicted to be activated or repressed by Rbfox, or exons for which the direction of Rbfox regulation cannot be determined unambiguously. In each group, exons are ranked by the confidence of prediction (left). (U)GCAUG motif scores, CLIP tag cluster scores, and inferred probability of Rbfox binding at different positions of the alternatively spliced region are shown in the grayscale heatmaps (darker colors represent stronger binding). UI5, E5, and DI5 represent regions near the 5′ splice sites of the upstream intron, exon, and downstream intron, respectively; Similarly, UI3, E3, and DI3 represent regions near the 3′ splice sites.

not overfit (Figure S3F), we predicted 772 cassette exons as direct Rbfox targets (FDR < 0.05; Table S6). Among these targets, Rbfox was predicted to activate 421 exons (probability of activation >0.7) and repress 113 exons (probability of repression >0.7), respectively. For the remaining 238 exons predicted as Rbfox targets, the Bayesian network was unable to assign the direction of regulation unambiguously (Figure 4F, left panel). This uncertainty is presumably due to a lack of observed Rbfox-dependent splicing in the current experimental settings and to binding of Rbfox in both upstream and downstream introns simultaneously (Figure 4F, right panel). Based on comparison of the predicted exons with previously validated Rbfox-regulated exons compiled from the literature, we estimated that our Bayesian network analysis has a sensitivity of 73%–79% (Supplemental Results; Figure S4; Table S7). We also compared the results of the Bayesian network analysis to our previous motif-based bioinformatic predictions and to another recent study that predicted Rbfox target exons based on the presence of the Rbfox motif sites and the correlation of exon splicing with Rbfox expression (Ray et al., 2013). These comparisons showed substantial overlap between exons predicted by different methods but also highlighted that the Bayesian network analysis effectively integrated features known to be consistent with regulated alternative splicing events, such as preservation of the reading frame and conservation of the alternative splicing pattern (Figure S5; Supplemental Results).

After we confirmed the performance of the Bayesian network, we applied the model to other types of alternative splicing events and predicted 212 events of tandem cassette exons (300 exons, Table S6) and 75 events of mutually exclusive exons (107 exons, Table S6) as direct Rbfox targets. Altogether, 587 genes have one or more alternative splicing events directly regulated by Rbfox. To understand the molecular function of these genes, we performed gene ontology (GO) analysis and found very significant enrichment of genes with annotated function in "cytoskeleton" (Benjamini FDR < 3.8×10^{-17}) and "neuron projection" (Benjamini FDR < 3.8×10^{-8}), compared to all brain-expressing genes (Table S8). In addition, proteins encoded by Rbfox target transcripts are enriched in PDZ domains that are known to be important for anchoring transmembrane proteins to the cytoskeleton and for functioning as scaffolds for signaling complexes (Benjamini FDR < 7.2×10^{-8}) (Ranganathan and Ross, 1997).

Rbfox Regulates Global Dynamic Splicing Changes during Brain Development

It has previously been shown that all three Rbfox family members undergo increased expression in the mouse brain at prenatal stages between embryonic day E12 and E18 (Tang et al.,

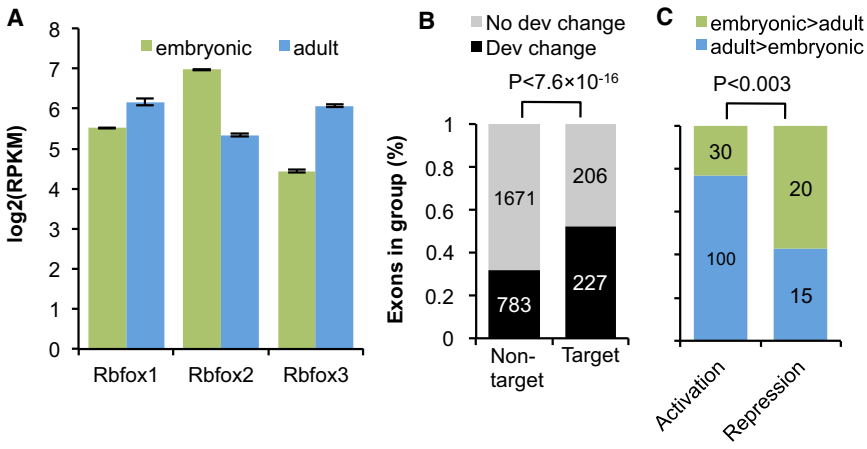

Figure 5. Rbfox Regulates Global Dynamic Splicing Changes during Brain Development

(A) Rbfox1, 2, and 3 expression in embryonic (green) or adult cortex (blue) as quantified by RNA-seq data. Error bars represent SEM.

(B) The proportion of Rbfox target exons and nontarget exons with developmental splicing changes. The difference is evaluated by Fisher's exact test.

(C) Rbfox target exons are divided into two groups depending on whether Rbfox activates or represses exon inclusion. For each group, the number of exons showing higher (blue) or lower (green) inclusion in the adult versus embryonic cortex is shown. The direction of developmental splicing change is compared with the direction of Rbfox regulation, as assessed by Fisher's exact test.

2009), and that the increase of Rbfox1 expression further extends into postnatal stages (Hammock and Levitt, 2011). In mouse and chicken, the differential expression of Rbfox proteins is correlated with splicing changes in several exons during CNS development (Kim et al., 2013; Tang et al., 2009), but how Rbfox proteins affect the global switch of the developmental splicing program is unclear.

We found that Rbfox family members undergo dynamic changes in expression between E17 and adult mouse cortex, as evaluated from a published RNA-seq data set (Dillman et al., 2013). Rbfox1 and Rbfox3 show 1.6-fold ($p < 0.02$; t test) and 3.1-fold ($p < 10^{-6}$; t test) increases in the adult cortex compared to E17 cortex, respectively, whereas the expression of Rbfox2 is reduced 3.1-fold ($p < 10^{-4}$; t test; Figure 5A). The expression changes of the Rbfox proteins parallel the splicing changes of Rbfox target exons: 55% of Rbfox target exons with sufficient read coverage to quantify splicing show splicing changes between the two developmental stages, as compared to 32% for exons not regulated by Rbfox (odds ratio = 2.4, $p < 7.6 \times 10^{-16}$; Fisher's exact test; Figure 5B). In addition, a majority (77%) of the exons activated by Rbfox have increased exon inclusion in the adult, whereas over half (57%) of the exons repressed by Rbfox have decreased exon inclusion (odds ratio = 4.4, $p < 0.003$; Fisher's exact test; Figure 5C). This asymmetry indicates that increased expression of Rbfox1 and Rbfox3 predominates the developmental splicing change of their targets, and in general they promote the switch to the adult splicing program through direct regulation.

Rbfox Target Genes Are Linked to Autism

We previously demonstrated that Rbfox target transcripts predicted bioinformatically based on conserved motif sites have significant overlap with genes implicated in autism, supporting the notion that disruption of either Rbfox1 itself or of its targets observed in autism patients is likely pathogenic (Zhang et al., 2010). This hypothesis was further supported by recent findings that *RBFOX1* is a hub in gene coexpression networks based on microarray profiling of autistic and control postmortem human brains, and its reduced expression in a subset of autism patients is correlated with altered splicing of predicted Rbfox target exons (Voineagu et al., 2011). The comprehensive Rbfox target network defined by integrative modeling now allows us to examine the link between *RBFOX1* and autism in more detail.

Among the 235 Rbfox target cassette exons that are conserved in human and that have sufficient RNA-seq read coverage to evaluate splicing change in autistic versus control brains (Voineagu et al., 2011), 97 (41%) show alteration of splicing in autistic brains ($|\Delta I| \geq 0.1$, and FDR ≤ 0.05), a very significant overlap compared to random chance (odds ratio = 3.4, $p < 1.4 \times 10^{-16}$; Fisher's exact test; Figure 6A). The autistic brains compared here were selected to have low expression level of Rbfox1 (Voineagu et al., 2011) (4.1-fold downregulation compared to control brains as measured by RNA-seq; $p < 0.05$, t test; Figure 6B). However, the massive splicing change of Rbfox targets detected in autistic versus control brains is somewhat surprising given the redundant role of the other Rbfox family members, as observed in *Rbfox1* KO mice (Gehman et al., 2011, 2012). Interestingly, we found that the expression of the other two family members, Rbfox2 and Rbfox3, was also downregulated 3.3-fold ($p < 0.05$; t test) and 3.2-fold ($p < 0.09$; t test), respectively. Therefore, simultaneous downregulation of all Rbfox family members might explain the massive splicing misregulation of Rbfox target exons observed in these autism patients. On the other hand, many of the splicing changes observed in autism patients may not be regulated by Rbfox proteins directly.

To focus on Rbfox target genes that are likely genetic risk factors of autism, we examined candidate autism-susceptibility genes in the SFARI autism gene database (Basu et al., 2009). Among the 519 candidate autism-susceptibility genes with mouse orthologs, 48 were identified as Rbfox targets by Bayesian network analysis (odds ratio = 2.8, $p < 9.6 \times 10^{-18}$, Fisher's exact test; Table 1 and Table S6). The list includes three genes that are currently regarded as causal in syndromic autism spectrum disorders (ASDs): *Shank3* (Phelan-McDermid Syndrome), *Cacna1c* (Timothy syndrome), and *Tsc2* (tuberous sclerosis complex). For a specific example, Rbfox is predicted to activate the inclusion of alternative exon 25 in the *Tsc2* gene, which is conserved between human and mouse (Figure 6C). Although the function of this alternative exon has not been characterized, its inclusion was recently shown to be

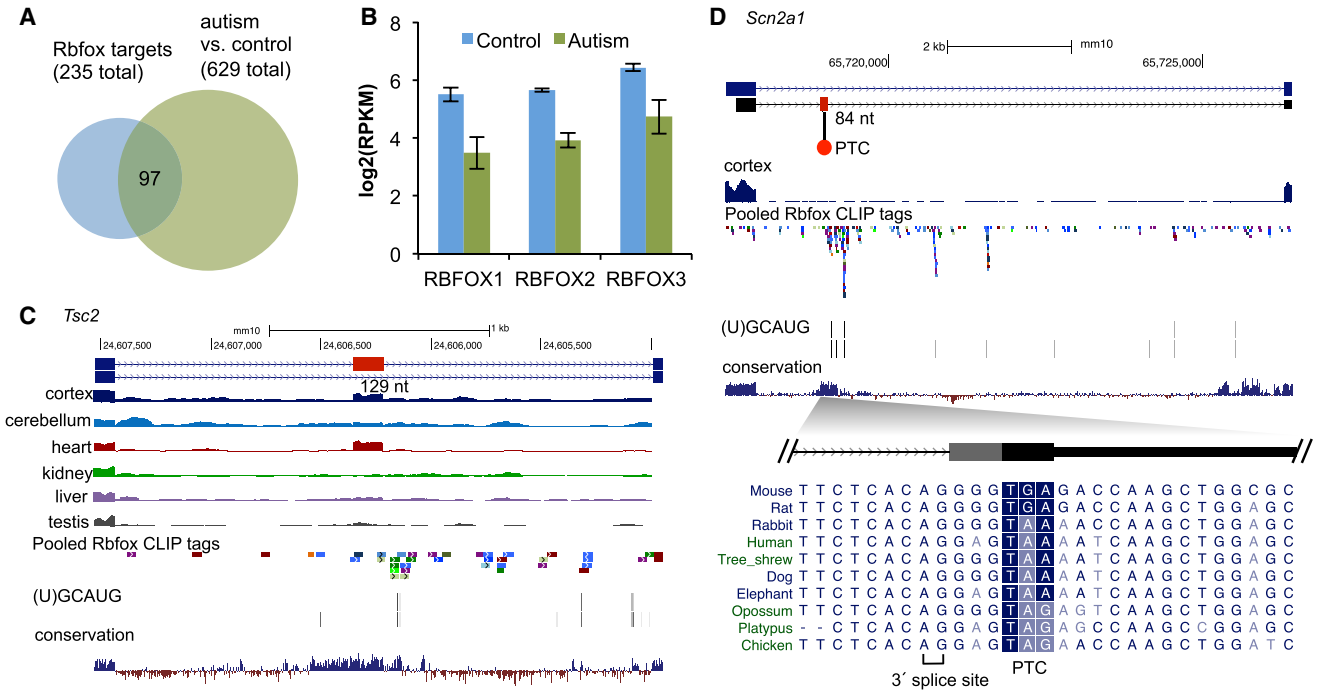

Figure 6. Rbfox Target Exons in Candidate Autism-Susceptibility Genes

(A) Overlap between Rbfox target cassette exons and exons with altered splicing in autistic versus control brains.

(B) Downregulation of Rbfox1, 2, and 3 expression in autistic versus control brains as quantified by RNA-seq. Error bars represent SEM.

(C) Rbfox is predicted to activate the inclusion of a 129 nt exon in the *Tsc2* gene. Below the gene structure schematic are RNA-seq data of different tissues showing a higher inclusion of the exon in cortex and heart, pooled Rbfox CLIP tags, Rbfox binding UGCAUG or GCAUG elements, and the phyloP conservation score.

(D) Rbfox is predicted to activate the inclusion of an 84 nt poisonous exon in the *Scn2a1* gene, which creates an in-frame premature termination codon (PTC) conserved in vertebrates.

dependent on Rbfox2 using cell culture models of epithelial-to-mesenchymal transition (Braeutigam et al., 2013).

Inclusion or exclusion of Rbfox target exons mostly introduces alteration in local amino acid sequences. However, we identified several cases (*Fat1*, *St7*, *Scn2a1*, and *Scn8a*) in which alternative splicing is potentially coupled with nonsense-mediated mRNA decay (NMD; Maquat, 2004). In *Scn2a1*, for instance, Rbfox is predicted to activate a cryptic exon harboring an in-frame premature stop codon via extensive binding to the downstream intron (Figure 6D). Inclusion of this exon is undetectable in the brain from RNA-seq data, presumably due to NMD of the inclusion isoform. Interestingly, whereas this alternative exon is conserved in vertebrates, compensatory mutations in the stop codon have accumulated. This has resulted in different stop codons in different species, suggesting an evolutionary selection pressure to preserve a stop codon. *Scn2a1* is one of the few genes with recurrent de novo gene-disrupting mutations according to exome sequencing of ASD patients (Ronemus et al., 2014), and our analysis suggests that the same gene can potentially be disrupted in autism by different mechanisms.

We tested a select set of exons in potentially autism-related genes for regulation by Rbfox1 in the mouse brain. The splicing level of these exons was measured in CNS-specific *Rbfox1*-KO mouse brains compared to WT controls by semiquantitative RT-PCR. Among eight exons for which both the inclusion and exclusion isoforms can be detected, three exons showed altered splicing upon Rbfox1 depletion ($|\Delta I| \geq 0.1$; Figure S6; Table S5), and these were not identified as Rbfox1 targets in the previous genome-wide analysis (Gehman et al., 2011). The relatively small number of validated exons is likely due to the redundancy of Rbfox proteins and the compensatory increase of Rbfox2 protein expression in *Rbfox1* KO animals (Gehman et al., 2011), although it is possible that some of the predicted targets might represent false-positives.

DISCUSSION

The focus of this study is to define and characterize the Rbfox target splicing-regulatory network in the mammalian brain. An important piece of information missing in previous efforts toward this aim (e.g., Fogel et al., 2012; Gehman et al., 2011, 2012; Ray et al., 2013; Zhang et al., 2008) is a genome-wide, high-resolution map of in vivo Rbfox interaction sites in the brain. Such a map is especially essential due to the functional redundancy of different Rbfox family members, so that simultaneous depletion of more than one member is probably required to uncover a majority of Rbfox-dependent exons in a physiologically relevant condition. A critical aspect of this work is our ability to identify

Table 1. List of Rbfox Target Genes Implicated in Autism

Entrez Gene ID (Mouse)	Gene Symbol	Tissue Expression	Autism versus Control Splicing Change	Function
11538	Adnp	non-brain-specific		transcription factor
11784	Apba2	brain-specific	Y	neurotransmitter release
11789	Apc	non-brain-specific		Wnt signaling
11941	Atp2b2	brain-specific		calmodulin binding
319974	Auts2	non-brain-specific		unknown
30948	Bin1	non-brain-specific	Y	synaptic vesicle endocytosis
12288	Cacna1c	non-brain-specific		ion channel
12289	Cacna1d	non-brain-specific		ion channel
12291	Cacna1g	non-brain-specific		ion channel
54725	Cadm1	non-brain-specific	Y	cell adhesion
320405	Cadps2	non-brain-specific	Y	synaptic vesicle exocytosis
100072	Camta1	non-brain-specific		calmodulin binding
67300	Cltc	non-brain-specific		receptor localization
104318	Csnk1d	non-brain-specific		cell signaling
74006	Dnm1l	non-brain-specific		mitochondrial and peroxisomal division
75560	Ep400	non-brain-specific		chromatin binding
13876	Erg	non-brain-specific		transcription factor
14107	Fat1	non-brain-specific		cell adhesion
268566	Gphn	non-brain-specific		receptor localization
74053	Grip1	non-brain-specific		glutamate receptor binding
108071	Grm5	brain-specific		G-protein coupled receptor
227753	Gsn	non-brain-specific	Y	cytoskeleton
14886	Gtf2i	non-brain-specific	Y	transcription factor
16531	Kcnma1	brain-specific	Y	ion channel
242274	Lrrc7	brain-specific		synapse assembly
18027	Nfia	non-brain-specific		transcription factor
319504	Nrcam	brain-specific	Y	cell adhesion
18189	Nrxn1	brain-specific		cell adhesion
18191	Nrxn3	brain-specific		cell adhesion
80883	Ntng1	brain-specific		cell adhesion
233977	Ppfia1	non-brain-specific	Y	cell signaling
353211	Prune2	non-brain-specific	Y	unknown
268859	Rbfox1	non-brain-specific		RNA metabolism
268902	Robo2	non-brain-specific		axon guidance
110876	Scn2a1	brain-specific		ion channel
20273	Scn8a	brain-specific		ion channel
58234	Shank3	non-brain-specific		synapse assembly
76376	Slc24a2	brain-specific		ion transport
72055	Slc38a10	non-brain-specific		ion/amino acid transport
94229	Slc4a10	brain-specific		ion transport
64213	St7	non-brain-specific		cell signaling
53416	Stk39	non-brain-specific		serine/threonine kinase
20910	Stxbp1	non-brain-specific	Y	neurotransmitter release
64009	Syne1	non-brain-specific	Y	cytoskeleton
22084	Tsc2	non-brain-specific		chaperone
22138	Ttn	low-brain-expression		sarcomere structure

(*Continued on next page*)

Table 1. *Continued*

Entrez Gene ID (Mouse)	Gene Symbol	Tissue Expression	Autism versus Control Splicing Change	Function
22214	Ube2h	non-brain-specific		protein ubiquitination
68134	Upf3b	non-brain-specific		mRNA metabolism

Tissue specificity of gene expression is based on the RNA-seq data set in Brawand et al. (2011), and splicing change in autistic versus control brains is based on the RNA-seq data in Voineagu et al. (2011). Gene function was compiled from the literature.

over 40,000 robust Rbfox binding sites using two HITS-CLIP protocols applied in parallel to each of the three Rbfox family members. These include 8,811 sites (2,298 CIMS and 6,606 CITS with 93 common sites) for which the exact site of protein-RNA crosslinking can be deduced, allowing for determination of protein-RNA interactions at a single nucleotide resolution. The overall distribution of Rbfox binding sites defined by HITS-CLIP is generally consistent with a recently published independent study (Lovci et al., 2013).

Detailed analysis of Rbfox-RNA crosslink sites provided insights into the biophysical principles of protein-RNA crosslinking. UV crosslinking of protein and RNA was thought to be most efficient for the nucleotide uridine. This presumptive preference was used to interpret the U stretch enriched in sequences around CIMS and in iCLIP data for several RBPs including Nova, Ago, hnRNP C, and TIA (Sugimoto et al., 2012), despite the fact that these proteins have a genuine binding preference for uridine. Therefore, the predominant crosslinking of Rbfox and substrate RNA at the two guanines in the (U)GCAUG motif, as suggested by CIMS and CITS analysis, is somewhat unexpected, especially given the presence of two uridines in the motif. We argue that these crosslink sites likely reflect the residues in closest contact with the protein or those interacting with specific amino acids, which agrees very well with a structure study of the protein-RNA complex (Auweter et al., 2006). Additional evidence supporting this argument comes from predominant crosslinking to different nucleotides for several other RBPs with distinct binding specificity (Moore et al., 2014). On the other hand, we also observed that a substantial proportion of Rbfox-RNA crosslink sites do not overlap with the canonical Rbfox (U)GCAUG motif. Paradoxically, when these sites were included for analysis, the base composition at CIMS identified in Rbfox CLIP data is slightly biased toward uridine. One possible interpretation for this discrepancy is that some of these sites without the Rbfox motif might have resulted from crosslinking of more transient protein-RNA interactions largely independent of the Rbfox specificity (e.g., recruitment by other interacting RBPs), which would suggest that preferential crosslinking to uridine indeed exists to some extent. However, for high-affinity protein-RNA interactions, such "background" preference does not prevent UV crosslinking at specific nucleotides in the core motif.

It has been reported that the rate of cDNA truncation at the crosslink site during reverse transcription is as high as 82% for Nova and over 95% for several other RBPs (Sugimoto et al., 2012). Based on the relative frequency of deletions in Rbfox CLIP tags obtained with standard and BrdU-CLIP protocols (14% versus 6.2%), we estimated that 57% of Rbfox CLIP tags are truncated at the crosslink sites (Experimental Procedures). Therefore, this parameter could vary substantially for different proteins, depending on the specific amino acid(s) and nucleotide(s) at the crosslink sites, and possibly also on different experimental conditions.

The second critical aspect of this study is the use of an integrative modeling approach to combine multiple complementary types of data, so that individually weak bits of information can be integrated to make strong predictions of Rbfox targets (Zhang et al., 2010). Because increasing amounts of high-throughput data are being produced for different RBPs, interrogating RNA regulation from different perspectives, such a method has the unique advantage of being able to identify direct, functional target transcripts with high specificity and sensitivity simultaneously. We were able to assign the direction of Rbfox regulation for a majority of Rbfox targets (69% of cassette exons), which allowed us to evaluate the impact of splicing regulation by Rbfox proteins in different physiological contexts, including brain development and autism.

Our analysis extends previous observations regarding the differential expression of Rbfox family members during brain development (Hammock and Levitt, 2011; Kim et al., 2013; Tang et al., 2009) by showing increased expression of Rbfox1 and Rbfox3 and decreased expression of Rbfox2 in adult compared to a late prenatal stage. Similar dynamic changes in Rbfox1 and Rbfox2 expression were previously observed in the heart during postnatal development (Kalsotra et al., 2008). Importantly, differential expression of Rbfox is paralleled by developmental splicing changes in over half of the quantifiable Rbfox target exons, frequently in the direction consistent with direct Rbfox regulation. Combined with neurodevelopmental defects observed in cell cultures (Kim et al., 2013), animal models (Gehman et al., 2011, 2012; Kim et al., 2013), and human patients (Bhalla et al., 2004) where Rbfox proteins are disrupted, this presents a compelling case for Rbfox proteins playing critical roles in driving the global dynamic change of the developmental splicing-regulatory program.

Given the strong implication of Rbfox1 in autism and the role of Rbfox proteins in neural development, we are particularly interested in the molecular mechanisms underlying ASD cases with mutations in *RBFOX1*. One hypothesis of the autism etiopathology is that many genes implicated in autism can be disrupted individually in *cis* by mutations in the genes themselves, or in *trans* by disruption of their upstream regulators such as Rbfox1. This hypothesis was supported both by the significant overlap between candidate autism-susceptibility genes and Rbfox target genes and by the significant overlap of predicted Rbfox target exons and exons altered in autistic brains with downregulation of Rbfox1 (Voineagu et al., 2011). Importantly,

we found that all Rbfox family members are downregulated in some autistic brains, underscoring the potential clinical relevance of the Rbfox target network in autism. In addition, we were able to highlight 48 genes in the SFARI autism gene database as direct Rbfox targets. The functions of these genes, which include cytoskeleton and scaffolding, synaptic transmission, ion channels, and transcription regulation, are potentially relevant to the neurobiology underlying autism.

We note that this study is aimed at defining the pan-Rbfox alternative splicing target network as a means of elucidating their molecular functions. We are currently limited in our ability to comprehensively identify target transcripts differentially regulated by the three family members individually. An outstanding question is the extent to which the Rbfox family members have distinct physiological functions and the underlying molecular mechanisms for these potential differences. Our CLIP data suggest that all three members have very similar protein-RNA interaction profiles, indicating that the highly conserved RRM is the major determinant of targeting specificity. However, the functional divergence of the different members could arise from the increased variation in the N-terminal and C-terminal regions, which were shown to also be important in splicing regulation (Jin et al., 2003; Nakahata and Kawamoto, 2005; Sun et al., 2012). For example, different members might recruit different cofactors and exert different regulatory effects depending on cellular context. However, addressing this question will require expression of different combinations of Rbfox proteins in physiologically relevant systems, a currently nontrivial experiment. Such data, when available, will nevertheless further improve the accuracy of our model by correlating them with protein-RNA interactions and other types of data specific to individual Rbfox family members.

Finally, another important question beyond the scope of this work concerns the roles of Rbfox in the regulation of other pathways in RNA metabolism. Several recent findings propose that Rbfox can also regulate alternative polyadenylation (Wang et al., 2008) and mRNA stability (Ray et al., 2013), based on analysis of Rbfox motif sites in 3′ UTRs that are correlated with transcript abundance in different tissues. Although the current data supporting the role of Rbfox in these processes are largely correlative, they are corroborated by our observation that almost 30% of Rbfox binding sites are in 3′ UTRs. A mechanistic understanding of the functional impact of Rbfox interacting with 3′ UTRs awaits further investigation. We expect that the data presented in this work will provide a valuable resource to facilitate these efforts.

EXPERIMENTAL PROCEDURES

All animal-related procedures were conducted according to the Institutional Animal Care and Use Committee guidelines at Columbia University Medical Center and the Rockefeller University.

HITS-CLIP Experiments Using Mouse Brain

Antibodies used in HITS-CLIP experiments for individual Rbfox members were tested to ensure that there is no or minimal cross-reactivity between family members and to optimize the signal-to-noise ratios of IP in the CLIP condition. Rbfox1, 2, and 3 CLIP experiments were performed with whole-brain tissue lysate of P15 CD1 mice using two different protocols. The standard CLIP libraries were prepared as described previously (Darnell, 2010; Licatalosi et al., 2008; Moore et al., 2014). Details of the BrdU-CLIP protocol are described in the Supplemental Experimental Procedures.

Bioinformatic analysis of CLIP data was performed essentially as described previously to obtain unique CLIP tags and define CLIP tag clusters and peaks (Charizanis et al., 2012; Moore et al., 2014; Zhang and Darnell, 2011). Mutations in unique CLIP tags were recorded for CIMS analysis as described previously (Moore et al., 2014; Zhang and Darnell, 2011). CITS analysis was performed using a computational method similar to that used to analyze iCLIP data with several modifications (König et al., 2010; Wang et al., 2010). We removed CLIP tags with deletions, which presumably had read through the crosslink sites and clustered the remaining tags based on their potential truncation sites, and then used a stringent statistical test to evaluate the significance of the observed truncation frequency.

RNA-Seq in HeLa Cells

HeLa cells were infected with murine-stem-cell-virus-expressing shRbfox2 or the empty retroviral vector (control) (Zhang et al., 2008). To reduce splicing precursors and intermediates, cells were fractionated to enrich cytoplasmic RNA before sequencing. The downstream analysis was performed as described previously (Charizanis et al., 2012) except that the proportional change of exon inclusion was estimated using the ASPIRE algorithm (Ule et al., 2005b).

RT-PCR Validation

Semiquantitative RT-PCR in HeLa cells was performed as described previously (Zhang et al., 2008). For RT-PCR validation, total RNA was extracted from brains of WT and homozygous *Rbfox1* KO mice. Candidate exons for validation were selected to cover the whole range of prediction confidence, avoiding exons with complex splicing patterns or very high or low inclusion levels.

Integrative Modeling Using Bayesian Network Analysis

We used our established Bayesian network framework with some modifications (Zhang et al., 2010). We considered differential splicing between WT and *Rbfox1* KO mouse brain, differential splicing between shRbfox2 and control HeLa cells, tissue-specific splicing of exons in cortex and cerebellum compared to other tissues (liver, kidney, and testis), and tissue-specific splicing of exons in heart compared to other tissues (liver, kidney, and testis). For motif sites, we considered both UGCAUG and VGCAUG elements, and measured their strength by their MECS score.

To train the model, we used 121 cassette exons compiled from the literature that have been previously validated as Rbfox targets in humans or mice and assigned them a class label (activation or repression). We also included unlabeled exons that showed relatively large or no changes in inclusion in exon-junction microarray or RNA-seq data. This resulted in 551 nonredundant exons, representing both targets and nontargets, to estimate model parameters. The trained Bayesian network model was applied to 16,034 cassette exons, as well as 4,074 events of tandem cassette exons and 960 events of mutually exclusive exons, to predict direct Rbfox targets.

Details of all experimental protocols and computational analyses are described in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

Rbfox HITS-CLIP and RNA-seq data were deposited to the NCBI Short Reads Archive under accession number SRP035321.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Results, Supplemental Experimental Procedures, six figures, and eight tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.02.005.

AUTHOR CONTRIBUTIONS

C.Z. and R.B.D. initiated this study. A.M. developed the BrdU-CLIP protocol. C.Z. optimized and performed Rbfox HITS-CLIP experiments in the R.B.D.

lab. S.S. and Z.Z. generated HeLa cell RNA-seq data. S.M.W.-V. and C.Z. performed bioinformatics analysis with assistance from C.X. Q.Y. and M.H. performed the mouse work and collected WT and *Rbfox1* KO brain samples. S.S. and Q.Y. performed experimental validation of Rbfox targets in HeLa cells and mouse brains, respectively. N.F. and P.A.S. produced the Rbfox3 polyclonal antibody and performed the initial analysis. M.Q.Z., A.R.K., R.B.D., and C.Z. supervised the study. S.M.W.-V., Q.Y., and C.Z. wrote the paper with input from all authors.

ACKNOWLEDGMENTS

The authors would like to thank Masato Yano for generating pRbfox3 plasmid, Joe Luna for methodological development assistance, Yuan Yuan and Michael Moore for helpful discussion, Scott Dewell for high-throughput sequencing, Melis Kayikci and Jernej Ule for ASPIRE analysis, Jinbiao Ma and Judith Kribelbauer for assistance with the interpretation of the Rbfox1-RNA complex structure data, and Judith Kribelbauer for assistance with the preparation of Figure 2G. This work was supported by grants from the National Institutes of Health (GM74688 and HG001696 to M.Q.Z., GM42699 to A.R.K., NS81706 and NS34389 to R.B.D., and R00GM95713 to C.Z.) and Simons Foundation Autism Research Initiative (240432 to R.B.D. and 297990 to C.Z.). R.B.D. is a Howard Hughes Medical Institute Investigator.

Received: September 22, 2013
Revised: November 30, 2013
Accepted: February 4, 2014
Published: March 6, 2014

REFERENCES

Auweter, S.D., Fasan, R., Reymond, L., Underwood, J.G., Black, D.L., Pitsch, S., and Allain, F.H.-T. (2006). Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. EMBO J. *25*, 163–173.

Baraniak, A.P., Chen, J.R., and Garcia-Blanco, M.A. (2006). Fox-2 mediates epithelial cell-specific fibroblast growth factor receptor 2 exon choice. Mol. Cell. Biol. *26*, 1209–1222.

Barash, Y., Calarco, J.A., Gao, W., Pan, Q., Wang, X., Shai, O., Blencowe, B.J., and Frey, B.J. (2010). Deciphering the splicing code. Nature *465*, 53–59.

Basu, S.N., Kollu, R., and Banerjee-Basu, S. (2009). AutDB: a gene reference resource for autism research. Nucleic Acids Res. *37* (Database issue), D832–D836.

Bhalla, K., Phillips, H.A., Crawford, J., McKenzie, O.L.D., 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. J. Hum. Genet. *49*, 308–311.

Braeutigam, C., Rago, L., Rolke, A., Waldmeier, L., Christofori, G., and Winter, J. (2013). The RNA-binding protein Rbfox2: an essential regulator of EMT-driven alternative splicing and a mediator of cellular invasion. Oncogene http://dx.doi.org/10.1038/onc.2013.50.

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csárdi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M., et al. (2011). The evolution of gene expression levels in mammalian organs. Nature *478*, 343–348.

Charizanis, K., Lee, K.-Y., Batra, R., Goodwin, M., Zhang, C., Yuan, Y., Shiue, L., Cline, M., Scotti, M.M., Xia, G., et al. (2012). Muscleblind-like 2-mediated alternative splicing in the developing brain and dysregulation in myotonic dystrophy. Neuron *75*, 437–450.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.

Damianov, A., and Black, D.L. (2010). Autoregulation of Fox protein expression to produce dominant negative splicing factors. RNA *16*, 405–416.

Darnell, R.B. (2010). HITS-CLIP: panoramic views of protein-RNA regulation in living cells. Wiley Interdiscip Rev. RNA *1*, 266–286.

Dillman, A.A., Hauser, D.N., Gibbs, J.R., Nalls, M.A., McCoy, M.K., Rudenko, I.N., Galter, D., and Cookson, M.R. (2013). mRNA expression, splicing and editing in the embryonic and adult mouse cerebral cortex. Nat. Neurosci. *16*, 499–506.

Dredge, B.K., and Darnell, R.B. (2003). Nova regulates GABA(A) receptor γ2 alternative splicing via a distal downstream UCAU-rich intronic splicing enhancer. Mol. Cell. Biol. *23*, 4687–4700.

Fogel, B.L., Wexler, E., Wahnich, A., Friedrich, T., Vijayendran, C., Gao, F., Parikshak, N., Konopka, G., and Geschwind, D.H. (2012). RBFOX1 regulates both splicing and transcriptional networks in human neuronal development. Hum. Mol. Genet. *21*, 4171–4186.

Gallagher, T.L., Arribere, J., Adkar, S., Marr, H., Dill, K., Garnett, A., Amacher, S., and Conboy, J. (2011). Fox1 and Fox4 regulate muscle-specific splicing in zebrafish and are required for cardiac and skeletal muscle functions. Dev. Biol. *344*, 491–492.

Gehman, L.T., Stoilov, P., Maguire, J., Damianov, A., Lin, C.-H., Shiue, L., Ares, M., Jr., Mody, I., and Black, D.L. (2011). The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. Nat. Genet. *43*, 706–711.

Gehman, L.T., Meera, P., Stoilov, P., Shiue, L., O'Brien, J.E., Meisler, M.H., Ares, M., Jr., Otis, T.S., and Black, D.L. (2012). The splicing regulator Rbfox2 is required for both cerebellar development and mature motor function. Genes Dev. *26*, 445–460.

Granneman, S., Kudla, G., Petfalski, E., and Tollervey, D. (2009). Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. Proc. Natl. Acad. Sci. USA *106*, 9613–9618.

Hammock, E.A.D., and Levitt, P. (2011). Developmental expression mapping of a gene implicated in multiple neurodevelopmental disorders, A2bp1 (Fox1). Dev. Neurosci. *33*, 64–74.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.

Jin, Y., Suzuki, H., Maegawa, S., Endo, H., Sugano, S., Hashimoto, K., Yasuda, K., and Inoue, K. (2003). A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. EMBO J. *22*, 905–912.

Kalsotra, A., Xiao, X., Ward, A.J., Castle, J.C., Johnson, J.M., Burge, C.B., and Cooper, T.A. (2008). A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. Proc. Natl. Acad. Sci. USA *105*, 20333–20338.

Kim, K.K., Nam, J., Mukouyama, Y.S., and Kawamoto, S. (2013). Rbfox3-regulated alternative splicing of Numb promotes neuronal differentiation during development. J. Cell Biol. *200*, 443–458.

König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat. Struct. Mol. Biol. *17*, 909–915.

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.

Lovci, M.T., Ghanem, D., Marr, H., Arnold, J., Gee, S., Parra, M., Liang, T.Y., Stark, T.J., Gehman, L.T., Hoon, S., et al. (2013). Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. Nat. Struct. Mol. Biol. *20*, 1434–1442.

Maquat, L.E. (2004). Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. Nat. Rev. Mol. Cell Biol. *5*, 89–99.

Martin, C.L., Duvall, J.A., Ilkin, Y., Simon, J.S., Arreaza, M.G., Wilkes, K., Alvarez-Retuerto, A., Whichello, A., Powell, C.M., Rao, K., et al. (2007). Cytogenetic and molecular characterization of A2BP1/FOX1 as a candidate gene for autism. Am. J. Med. Genet. B. Neuropsychiatr. Genet. *144B*, 869–876.

Minovitsky, S., Gee, S.L., Schokrpur, S., Dubchak, I., and Conboy, J.G. (2005). The splicing regulatory element, UGCAUG, is phylogenetically and spatially

conserved in introns that flank tissue-specific alternative exons. Nucleic Acids Res. *33*, 714–724.

Moore, M.J., Zhang, C., Gantman, E.C., Mele, A., Darnell, J.C., and Darnell, R.B. (2014). Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. Nat. Protoc. *9*, 263–293.

Nakahata, S., and Kawamoto, S. (2005). Tissue-dependent isoforms of mammalian Fox-1 homologs are associated with tissue-specific splicing activities. Nucleic Acids Res. *33*, 2078–2089.

Ponthier, J.L., Schluepen, C., Chen, W., Lersch, R.A., Gee, S.L., Hou, V.C., Lo, A.J., Short, S.A., Chasis, J.A., Winkelmann, J.C., and Conboy, J.G. (2006). Fox-2 splicing factor binds to a conserved intron motif to promote inclusion of protein 4.1R alternative exon 16. J. Biol. Chem. *281*, 12468–12474.

Ranganathan, R., and Ross, E.M. (1997). PDZ domain proteins: scaffolds for signaling complexes. Curr. Biol. *7*, R770–R773.

Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., et al. (2013). A compendium of RNA-binding motifs for decoding gene regulation. Nature *499*, 172–177.

Ronemus, M., Iossifov, I., Levy, D., and Wigler, M. (2014). The role of de novo mutations in the genetics of autism spectrum disorders. Nat. Rev. Genet. *15*, 133–141.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., et al. (2007). Strong association of de novo copy number mutations with autism. Science *316*, 445–449.

Sugimoto, Y., König, J., Hussain, S., Zupan, B., Curk, T., Frye, M., and Ule, J. (2012). Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. Genome Biol. *13*, R67.

Sun, S., Zhang, Z., Fregoso, O., and Krainer, A.R. (2012). Mechanisms of activation and repression by the alternative splicing factors RBFOX1/2. RNA *18*, 274–283.

Tang, Z.Z., Zheng, S., Nikolic, J., and Black, D.L. (2009). Developmental control of CaV1.2 L-type calcium channel splicing by Fox proteins. Mol. Cell. Biol. *29*, 4757–4765.

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

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

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.

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.

Voineagu, I., Wang, X., Johnston, P., Lowe, J.K., Tian, Y., Horvath, S., Mill, J., Cantor, R.M., Blencowe, B.J., and Geschwind, D.H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature *474*, 380–384.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470–476.

Wang, Z., Kayikci, M., Briese, M., Zarnack, K., Luscombe, N.M., Rot, G., Zupan, B., Curk, T., and Ule, J. (2010). iCLIP predicts the dual splicing effects of TIA-RNA interactions. PLoS Biol. *8*, e1000530.

Xu, B., Roos, J.L., Levy, S., van Rensburg, E.J., Gogos, J.A., and Karayiorgou, M. (2008). Strong association of de novo copy number mutations with sporadic schizophrenia. Nat. Genet. *40*, 880–885.

Yeo, G.W., Coufal, N.G., Liang, T.Y., Peng, G.E., Fu, X.-D., and Gage, F.H. (2009). An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nat. Struct. Mol. Biol. *16*, 130–137.

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.

Zhang, C., Zhang, Z., Castle, J., Sun, S., Johnson, J., Krainer, A.R., and Zhang, M.Q. (2008). Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. Genes Dev. *22*, 2550–2563.

Zhang, C., Frias, M.A., Mele, A., Ruggiu, M., Eom, T., Marney, C.B., Wang, H., Licatalosi, D.D., Fak, J.J., and Darnell, R.B. (2010). Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. Science *329*, 439–443.

Zhang, C., Lee, K.-Y., Swanson, M.S., and Darnell, R.B. (2013). Prediction of clustered RNA-binding protein motif sites in the mammalian genome. Nucleic Acids Res. *41*, 6793–6807.